

MONICA MENDES KANGUSSU MARCOLINO

CARACTERIZAÇÃO DAS NOVAS SUBFAMÍLIAS DE AMASTINAS E
DESENVOLVIMENTO DE TECNOLOGIA PARA GERAÇÃO DE NOCAUTES
CONDICIONAIS EM *Trypanosoma cruzi*

CURITIBA

2014

MONICA MENDES KANGUSSU MARCOLINO

CARACTERIZAÇÃO DAS NOVAS SUBFAMÍLIAS DE AMASTINAS E
DESENVOLVIMENTO DE TECNOLOGIA PARA GERAÇÃO DE NOCAUTES
CONDICIONAIS EM *Trypanosoma cruzi*

Tese apresentada ao Curso de Pós-Graduação em Ciências-Bioquímica, Departamento de Bioquímica e Biologia Molecular, Setor de Ciências Biológicas, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Doutor em Ciências-Bioquímica.

Orientador: Prof. Dr. Wanderson Duarte da Rocha

CURITIBA

2014

TERMO DE APROVAÇÃO

MONICA MENDES KANGUSSU MARCOLINO

CARACTERIZAÇÃO DAS NOVAS SUBFAMÍLIAS DE AMASTINAS E DESENVOLVIMENTO DE TECNOLOGIA PARA GERAÇÃO DE NOCAUTES CONDICIONAIS EM *Trypanosoma cruzi*.

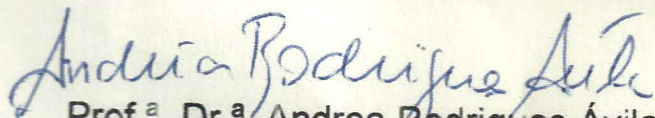
Tese aprovada com requisito parcial para obtenção do grau de Doutor no
Curso de Pós-Graduação em Ciências-Bioquímicas, Setor de Ciências
Biológicas da Universidade Federal do Paraná, pela seguinte banca
examinadora:


Orientador: Prof. Dr. Wanderson Duarte da Rocha

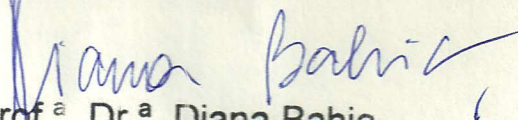
Departamento de Bioquímica e Biologia Molecular, UFPR


Prof. Dr. Marcelo Müller dos Santos

Departamento de Bioquímica e Biologia Molecular, UFPR


Prof^a. Dr^a. Andrea Rodrigues Ávila

Fundação Oswaldo Cruz (FIOCRUZ/PR) -- Instituto Carlos Chagas (ICC)


Prof^a. Dr^a. Diana Bahia

Departamento de Biologia Geral, UFMG


Prof. Dr. Edmundo Carlos Grisard

Centro de Ciências Biológicas, UFSC

Curitiba, 23 de abril de 2014

DEDICATÓRIA

Dedico aos meus avós e padrinhos Elena e Rui, que agora estão separados porque alguém no céu precisava de um parceiro na sinuca.

AGRADECIMENTOS

Agradeço a todos que colaboraram para a realização deste trabalho.

Primeiramente, agradeço ao Prof. Wanderson daRocha, por ser um excelente orientador, que se dedica à ciência de forma muito admirável. A grande evolução que obtive para minha formação como pesquisadora deve-se, em grande parte, a suas sugestões, conselhos e exigências, que me impulsionaram a desenvolver senso crítico, entre outras habilidades que ajudarão imensamente no decorrer da minha carreira. Obrigada por sua dedicação e exemplo.

Agradeço também ao Prof. Emanuel pelo grande apoio na minha entrada e primeiro ano de doutorado e aos membros da banca interna, Prof. Marcelo Müller e Prof. Stenio Fragoso pelas valiosas correções, sugestões e acompanhamento do trabalho durante seu desenvolvimento ao longo destes quatro anos.

Aos professores do Núcleo de Fixação de Nitrogênio (NFN) por permitirem o uso de instalações e equipamentos que foram essenciais para o andamento do trabalho. Ainda em relação ao NFN, agradeço aos técnicos, Valter Baura, Roseli Prado e Alexsandro Albani, pelo precioso trabalho. Também agradeço aos colegas, em especial ao Adriano Stefanello, Marco Aurélio Oliveira, Thiago Estefano e Vânia Pankiewicz pela gentileza em auxiliar no uso de equipamentos e no empréstimo de reagentes.

Aos professores e colegas do grupo de Oxidações Biológicas e de Química de Carboidratos pelo empréstimo de equipamentos e reagentes que foram de grande utilidade durante o desenvolvimento do projeto.

Aos pesquisadores do Instituto Carlos Chagas (ICC), especialmente ao Stenio Fragoso, Andréa Ávila e Bruno Dallagiovanna, pelo uso das instalações e equipamentos em diversos experimentos realizados no instituto. Agradeço também aos colegas Mariana Serpeloni, Eloise Guerra-Slompo, Daniella Gradia e Carla Lima pelo auxílio durante o desenvolvimento destes experimentos.

Ao Prof. Edivaldo Trindade pela solicitude para o uso do microscópio confocal e citômetro de fluxo. Agradeço também aos técnicos, Israel Bini, Alessandra Santos e Inês Rabitto pela atenção no uso destes equipamentos.

Ao Prof. Renato Mortara, UNIFESP - São Paulo, pela colaboração no trabalho e por me acolher em seu laboratório para realização de experimentos. Agradeço ao apoio de seus alunos Mario Cruz e Pilar Veras.

Agradeço a Prof.^a Santuza Teixeira, UFMG - Belo Horizonte, pela ampla colaboração no trabalho, principalmente em relação ao estudo das amastinas, no qual teve participação. Agradeço também ao pessoal do laboratório que me apoiou durante minha estadia, especialmente a Gabriela Burle, Viviane Silva e Rita Cardoso.

À Prof.^a Isabel Roditi por me acolher na Suíça e em seu laboratório e me orientar no desenvolvimento do trabalho com as Nucleases Dedo de Zinco. Foi um grande prazer e uma experiência incrível trabalhar e morar na Suíça, conhecendo pessoas e culturas diferentes. Agradeço o apoio de todos no laboratório, em especial a Gaby Schumann-Burkard, Mirella Woodert e Sebastian Künsel. Agradeço também aos amigos que fiz, que me ajudaram na vida fora do laboratório, em especial a Rishita Nag e Mariana Hidalgo.

Ao Dr. Jean-Paul Hermann por ceder os plasmídeos contendo as sequências de Cre59 e Cre60, utilizados na construção do plasmídio para expressão de DiCre recombinase em *T. cruzi*, utilizado neste trabalho. Agradeço também ao Dr. George Cross por ceder os plasmídeos pyrFEKO em diversas versões com genes de resistências diferentes flanqueadas por *loxP*.

Agradeço, com muito amor, aos colegas do laboratório GFP, Ana Paula Cunha pela parceria no trabalho com a DiCre; Geison Cambri, por me ensinar várias coisas no começo do trabalho e pelas ótimas dicas musicais e esportivas; Laiane Lemos, pela parceria em várias etapas do trabalho, por eu poder aprender a ensinar com uma aluna tão dedicada; Letícia Laibida, pelas altas discussões sobre protocolos, músicas, filmes e seriados; Normanda Souza Melo, pela parceria em

milhares de experimentos, protocolos padronizados, almoços, escaladas, e como se não bastasse, nos finais de semana também; e Rodrigo Soares por comprar várias lâmpadas e pela companhia nos cafés e escaladas. A convivência com todos vocês fez do dia a dia uma experiência divertida e cheia de aprendizado e me fez crescer muito, obrigada por adoçarem minha vida!

Aos colegas do departamento, pelo companheirismo, pelos churrascos e confraternizações. É muito bom trabalhar em um ambiente onde podemos fazer amigos e contar com as pessoas. Agradeço especialmente a Vânia, Marina, Norminha, Karina e Liziane. Aos demais, pessoal do cafezinho, da turma do mestrado, das conversas aleatórias, eu mando um abraço, são muitos para serem citados aqui.

À pós-graduação por todo trabalho que tiveram comigo, e por eu ter participado como representante discente nas reuniões do colegiado, agradeço ao Prof. Miguel Nosedá, Prof.^a Silvia Cadena e Irene Rosa.

A toda minha família, principalmente meus pais, Tadeu e Cláudia, e meus irmãos, Tadeu e André, por serem a base da minha educação e formação pessoal. Obrigada por todo o apoio e por me ensinarem a ser gente.

Aos amigos do coração, Mariana, Dalis, Rodrigo, Aninha, Willian e Willian S, por alegrarem meus finais de semana e me tirarem um pouco de dentro da minha própria cabeça. Vai um abraço especial pro Douglas Guisi que cuidou da nossa casa enquanto o Elio foi me visitar na Suíça.

Ao meu marido, Elio Marcolino, por ser meu porto seguro, onde amarro o barquinho das minhas confusões e consigo um pouco de calma. Obrigada por ser meu parceiro, dividir tua vida comigo, me divertir, me encorajar, me apoiar, cuidar dos nossos gatinhos e de mim quando eu fico doente. Amo você!

Por fim, agradeço a Deus por minha vida e pelo mundo ser um lugar tão interessante, cheio de desafios que fazem da nossa caminhada um eterno aprendizado. Obrigada!

EPÍGRAFE

*“Creation may be spiritual in origin, but that doesn't mean that everything created is spiritual. How can I explain such things to you? Let us accept the world is a mystery. Nature is neither solely material nor entirely spiritual. Man, too, is more than flesh and blood; otherwise, no religions would have been possible. Behind each cause is still another cause; the end or the beginning of all causes has yet to be found. **Yet, only one thing must be remembered: there is no effect without a cause, and there is no lawlessness in creation**”*

Albert Einstein, em Einstein and the Poet (1983)

RESUMO

O protozoário *Trypanosoma cruzi*, causador da doença de Chagas, é um parasita intracelular obrigatório. Logo, o melhor entendimento da função de proteínas envolvidas diretamente na interação parasito::hospedeiro poderá fornecer informações chaves no desenvolvimento de fármacos mais eficientes. As amastinas, inicialmente descritas como uma família de glicoproteínas de membrana mais expressas em formas amastigotas, apresentam alta variabilidade podendo ser estratificada em quatro subfamílias. Contudo, pouco se sabe sobre a função destas proteínas, já que estudos de genômica funcional em *T. cruzi* são limitados devido às poucas opções de ferramentas de manipulação da expressão. Neste contexto, uma boa opção seria o uso do sistema DiCre, que permite a deleção de longas regiões genômicas ou de marcas de seleção por recombinação sítio específica mediada pela recombinase Cre, cuja atividade é regulada pelo ligante rapamicina. Desta forma, este trabalho tem por objetivo a caracterização funcional das novas subfamílias de amastinas de *T. cruzi* por métodos convencionais, além da transferência de um sistema de nocaute condicional mediada pela recombinase DiCre. Como principais resultados, temos a reclassificação das amastinas de *T. cruzi* em δ -, δ -Ama40, β 1- e β 2-amastinas, sendo que as δ -amastinas possuem maior número de cópias (~14) e em múltiplos cromossomos no genoma. Para várias cepas testadas, a expressão de mRNA das δ -amastinas e δ -Ama40 é maior em amastigotas, enquanto que as β -amastinas em epimastigotas. Apesar de todas as amastinas estarem associadas a frações de membrana, elas parecem estar localizadas em compartimentos distintos, com alguma provável sobreposição. Paralelamente a estes esforços foram gerados vetores para utilização do sistema de nocautes condicionais utilizando DiCre em *T. cruzi* e sua atividade foi testada. Dados de *Southern blot* sugerem que a recombinação ocorreu de forma controlada após longo período de exposição ao ligante. Sendo assim, esta ferramenta poderá ser utilizada para a remoção de marcas de seleção, e excisão de longos trechos de DNA. Esta reduzida atividade de DiCre pode ser considerada positiva, uma vez que a expressão não regulada da recombinase CRE tem se mostrado tóxica diversos organismos.

Palavras-chave: *T. cruzi*, amastinas, subfamílias, sistema DiCre, recombinação.

ABSTRACT

Trypanosoma cruzi, the etiological agent of Chagas' disease, is an obligate intracellular parasite. Therefore, the understanding of protein function directly involved in host-parasite interaction might bring key information for new strategies of drug development. Amastins, initially described as amastigote-specific membrane glycoprotein family, present high variability allowing its classification into four subfamilies. Nevertheless, little is known regarding the amastin function, once functional genomics studies in *T. cruzi* are restricted due to the few options available for reverse genetics. In this scenario, one versatile option would be the usage of Di_Cre recombinase system. It is a site-specific recombinase activated by ligand, which catalyzes the recombination between two 34 bp sequences called LoxP permitting the deletion of long stretches of genomic DNA or drug-resistance marks. In this sense, here we aim to characterize the function of new *T. cruzi* amastin subfamilies by conventional methods beyond the transference of Di_Cre mediated conditional gene knockout system. As leading results, *T. cruzi* amastins were reclassified into δ -, δ -Ama40, β 1- e β 2-amastins. δ -amastins have greater number of copies (~14) and are present in multiple chromosomes in the genome. Expression of δ -amastins and δ -Ama40 mRNA is more abundant in amastigote forms, while β -amastins in epimastigotes. Although all amastins are associated to membrane fraction, they seem to be localized in distinct cell compartments with some overlapping. Alongside, plasmids to the DiCre mediated conditional gene knockout system were built to allow its expression in *T. cruzi*, and its activity was tested. Southern blot data suggest the recombination occurred in a controlled fashion after the ligand exposure for longer periods. Therefore, this tool may be used to the excision of drug-resistance marks and/or long genomic DNA sequences. The reduced activity of Di_Cre may be considered positive result since Cre recombinase expression has been detected as toxic for several organisms.

Key words: *T. cruzi*, amastins, subfamilies, Di_Cre system, recombination.

LISTA DE FIGURAS

FIGURA 1 – CICLO DE VIDA DO <i>Trypanosoma cruzi</i>	21
FIGURA 2 - ESQUEMA REPRESENTATIVO DA EXPRESSÃO GÊNICA EM TRIPANOSSOMATÍDEOS	24
FIGURA 3 – TOPOLOGIA PREDITA DAS PROTEÍNAS DE SUPERFÍCIE AMASTINAS.....	30
FIGURA 4 – ORGANIZAÇÃO GENÔMICA DAS ISOFORMAS DAS AMASTINAS DE <i>T. cruzi</i>	32
FIGURA 5 – EFEITO DE RECOMBINAÇÃO SÍTIO-ESPECÍFICA	38
FIGURA 6 – ESQUEMA REPRESENTATIVO DO SISTEMA DICRE.....	41
FIGURA 7 – ESQUEMA DE CONSTRUÇÃO DE pROCK FEKO_Pur Neo	48
FIGURA 8 – ESQUEMA DE CONSTRUÇÃO DO PLASMÍDIO pODiCre	49
FIGURA 9 – ANÁLISE FILOGENÉTICA DAS SEQUÊNCIAS DE AMASTINAS DE DIFERENTES CEPAS DE <i>T. cruzi</i>	69
FIGURA 10 – CONSERVAÇÃO DA SEQUÊNCIA DE AMINOÁCIDOS DAS SUBFAMÍLIAS DE AMASTINAS	70
FIGURA 11 – LOCALIZAÇÃO GENÔMICA DOS GENES DE AMASTINAS PARA DIFERENTES CEPAS/CLONES DE <i>T. cruzi</i>	74
FIGURA 12 – EXPRESSÃO DO mRNA DE AMASTINAS DURANTE O CICLO BIOLÓGICO DE <i>T. cruzi</i> EM DIFERENTES CEPAS/CLONES.....	78
FIGURA 13 – PREDIÇÃO DA TOPOLOGIA DAS SUBFAMÍLIAS DE AMASTINAS.....	80
FIGURA 14 – LOCALIZAÇÃO SUBCELULAR DE DIFERENTES AMASTINAS EM FUSÃO COM GFP.....	83
FIGURA 15 – LOCALIZAÇÃO SUBCELULAR DE δ -AMA-40 EM FUSÃO COM GFP.....	83
FIGURA 16 – DISTRIBUIÇÃO DAS PROTEÍNAS AMASTINAS EM FRAÇÕES CELULARES.....	84
FIGURA 17 – EFEITO DA EXPRESSÃO DE AMASTINAS FUSIONADAS GFP NO CRESCIMENTO DE EPIMASTIGOTAS.	87

FIGURA 18 – INFLUÊNCIA DO CRESCIMENTO DE FORMAS EPIMASTIGOTAS NA EXPRESSÃO DAS PROTEÍNAS DE FUSÃO A GFP	89
FIGURA 19 – ESQUEMA REPRESENTATIVO DA ESTRATÉGIA DE CONSTRUÇÃO DO VETOR PARA DELEÇÃO DE δ -Ama40 e β -AMASTINAS POR RECOMBINAÇÃO HOMÓLOGA.....	91
FIGURA 20 – ANÁLISE DA INTEGRAÇÃO DOS CASSETES NO GENOMA DE <i>T. cruzi</i>	93
FIGURA 21 – ANÁLISE DA INTEGRAÇÃO DO CASSETE UP32-Neo-DN32 NO LOCUS DAS β -AMASTINAS PRESENTES NO CROMOSSOMO 32	94
FIGURA 22 – SENSIBILIDADE DE FORMAS EPIMASTIGOTAS DE <i>T.cruzi</i> A PUROMICINA	98
FIGURA 23 – PERFIL ELETROFORÉTICO POR PFGE E DE HIBRIDIZAÇÃO DE DNA CROMOSSOMAL COM SONDA RADIOATIVA	99
FIGURA 24 – ESQUEMA REPRESENTATIVO DO VETOR pROCK FEKO_Pur Neo ANTES E APÓS RECOMBINAÇÃO MEDIADA POR CRE RECOMBINASE.....	102
FIGURA 25 – SOUTHERN BLOT DO TESTE DE RECOMBINAÇÃO MEDIADO POR CRE RECOMBINASE INDUZIDA POR RAPAMICINA	103
FIGURA 26 – ANÁLISE POR SOUTHERN BLOT DO TESTE DE RECOMBINAÇÃO MEDIADO POR CRE RECOMBINASE INDUZIDA POR RAPAMICINA EM DIFERENTES CONCENTRAÇÃO E TEMPOS DE INDUÇÃO..	104
FIGURA 27 – TESTE DE RECOMBINAÇÃO MEDIADO POR DICRE RECOMBINASE APÓS CLONAGEM.	106

LISTA DE SIGLAS E SÍMBOLOS

α - ³² P-dCTP	2'-desoxicitidina-5'trifosfato marcada com fósforo 32
AE	amastigota extracelular
cDNA	DNA complementar
CRE	recombinase Cre
DAPI	4',6-diamidino-2-phenylindole,
DiCre	recombinase CRE Dimerizável
DNA	ácido desoxiribonucléico
<i>floxed</i>	fragmento de DNA flanqueado por sítios <i>loxP</i>
G418	geneticina
GFP	proteína verde fluorescente
gGAPDH IR	Região intergênica de GAPDH
gp63	glicoproteína 63
gp72	glicoproteína 72
GPI	glicofosfatidilinositol
gRNA	RNA guia
Higro	higromicina
HIV	vírus de imunodeficiência humana
HSVTK	timidina quinase do vírus da <i>Herpes simplex</i>
HX1	região 5'UTR da proteína ribossômico TcP2β
LA	meio de cultivo de bactérias Luria Agar
LB	meio de cultivo de bactérias Luria Broth
<i>loxP</i>	sítio de recombinação para recombinase CRE
LTR	transposon com regiões terminais longas

MASP	proteínas associadas a mucinas
mRNA	RNA mensageiro
Neo	neomicina
ORF	fase aberta de leitura
PAC	resistência a puromicina
pb	pares de bases
PBS	tampão salino-fosfato
PCR	reação em cadeia da polimerase
PEPCK	fosfoenolpiruvato carboxiquinase
PFGE	eletroforese de campo pulsátil
RDE	representação da expressão diferencial
RFLP	polimorfismo do comprimento do fragmento de restrição
RLUC	renila luciferase
RNA	ácido ribonucléico
RNA pol I	RNA polimerase I
RNA pol II	RNA polimerase II
RNAi	RNA de interferência
rRNA	RNA ribossômico
SFB	soro fetal bovino
SSC	tampão citrato de sódio/ cloreto de sódio
SSP4	proteína específica de superfície - 4
TCT	tripomastigota de cultura em tecido
TetR	repressor de tetraciclina
U	unidade
UTR	região não traduzida

WT	selvagem
ZFN	nuclease dedo de zinco
α -GFP	anticorpo anti- proteína verde fluorescente
α -gp63	anticorpo anti-glicoproteína 63

SUMÁRIO

1	INTRODUÇÃO.....	17
2	REVISÃO BIBLIOGRÁFICA	19
2.1	DOENÇA DE CHAGAS.....	19
2.2	CICLO BIOLÓGICO DO PARASITO.....	20
2.3	EXPRESSÃO GÊNICA EM TRIPANOSSOMATÍDEOS.....	23
2.4	GENES DIFERENCIALMENTE EXPRESSOS DURANTE O CICLO BIOLÓGICO DE <i>T. CRUZI</i>	25
2.5	AMASTINAS	29
2.6	FORMAS DE MANIPULAÇÃO DA EXPRESSÃO GÊNICA EM TRIPANOSSOMATÍDEOS.....	34
2.7	O SISTEMA CRE-LOXP E DICRE-LOXP	37
3	JUSTIFICATIVA	43
4	OBJETIVOS.....	45
4.1	OBJETIVOS ESPECÍFICOS:.....	45
5	MATERIAIS E MÉTODOS.....	46
5.1	ANÁLISE DE SEQUÊNCIAS DO GENOMA	46
5.2	CULTIVO	46
5.2.1	Bactérias	46
5.2.2	Parasitos	47
5.3	VETORES.....	47
5.3.1	Construídos anteriormente	47
5.3.2	Outros vetores	50
5.4	INICIADORES.....	51
5.5	AMPLIFICAÇÃO DO DNA POR PCR	52
5.6	RESTRIÇÃO ENZIMÁTICA DE DNA	52
5.7	ELETROFORESE DE DNA	53
5.8	PURIFICAÇÃO DE FRAGMENTOS	53
5.9	REAÇÕES DE LIGAÇÃO.....	54

5.10	PREPARO DE CÉLULAS TERMOCOMPETENTES E TRANSFORMAÇÃO BACTERIANA POR CHOQUE TÉRMICO	54
5.11	OBTENÇÃO DE DNA PLASMIDIANO	55
5.12	EXTRAÇÃO DE DNA GENÔMICO	56
5.13	TRANSFEÇÃO DE PARASITOS.....	56
5.14	CLONAGEM DOS PARASITOS TRANSFECTADOS.....	57
5.15	CURVA DE RESISTÊNCIA A PUROMICINA	57
5.16	TESTE DE RECOMBINAÇÃO POR RAPAMICINA	58
5.17	OBTENÇÃO DE EXTRATO PROTÉICO TOTAL DE EPIMASTIGOTAS DE <i>T. CRUZI</i>	59
5.18	OBTENÇÃO DA FRAÇÃO ENRIQUECIDA DE MEMBRANA DE EPIMASTIGOTAS DE <i>T. CRUZI</i>	59
5.19	SDS-PAGE DO EXTRATO PROTÉICO TOTAL E DAS FRAÇÕES ENRIQUECIDAS	60
5.19.1	Coloração com Azul de Comassie	60
5.19.2	Coloração com Prata	60
5.20	WESTERN BLOT DE PROTEÍNAS DE EPIMASTIGOTAS DE <i>T. CRUZI</i>	61
5.21	ELETROFORESE DE CAMPO PULSADO (PFGE).....	61
5.22	TESTE DE POLIMORFISMO DO COMPRIMENTO DO FRAGMENTO DE RESTRIÇÃO (RFLP) E SOUTHERN BLOT.....	63
5.23	NORTHERN BLOT	63
5.24	MARCAÇÃO DE SONDAS RADIOATIVAS E HIBRIDIZAÇÃO	64
5.25	MICROSCOPIA CONFOCAL E IMUNOFLUORESCÊNCIA.....	65
5.26	CURVA DE CRESCIMENTO	66
5.27	CITOMETRIA DE FLUXO	66
6	RESULTADOS E DISCUSSÃO.....	67
6.1	PARTE 1: CARACTERIZAÇÃO FUNCIONAL DAS NOVAS SUBFAMÍLIAS DE AMASTINAS	67
6.1.1	Conservação, estrutura e organização genômica das subfamílias de amastinas de <i>T. cruzi</i>	67
6.1.2	Expressão das subfamílias de amastinas de <i>T. cruzi</i>	75

6.1.3	Localização subcelular das amastinas de <i>T. cruzi</i>	78
6.1.4	Deleção dos genes que codificam isoformas das proteínas amastinas de <i>Trypanosoma cruzi</i>	89
6.2	PARTE 2: DESENVOLVIMENTO DO SISTEMA DICKRE PARA OBTENÇÃO DE NOCAUTES CONDICIONAIS EM <i>TRYPANOSOMA CRUZI</i>	96
6.2.1	Teste de recombinação por DiCre recombinase	100
7	CONCLUSÃO	108
8	PERSPECTIVAS	110
9	REFERÊNCIAS BIBLIOGRÁFICAS.....	111
ANEXO 1	120
ANEXO 2	137

1 INTRODUÇÃO

A doença de Chagas é uma endemia da América Latina, que já se espalhou pelo mundo devido aos fluxos migratórios e hoje acomete de 8 a 10 milhões de indivíduos. As opções de tratamento disponíveis apresentam fortes efeitos colaterais e parasitos resistentes, o que evidencia a importância de novas alternativas para este cenário. Neste contexto, o estudo de proteínas diferencialmente expressas durante o ciclo biológico é interessante já que estas podem ter uma função na diferenciação ou adaptação do parasito no ambiente em que está vivendo. Sendo assim, interferir em sua função poderia culminar com a morte ou menor proliferação do parasito. Além disso, respostas imunológicas geradas contra estas moléculas poderão favorecer o hospedeiro, tornando-o capaz de eliminar este parasito ou de reduzir danos causados pela infecção.

Uma família de proteínas diferencialmente expressas foi inicialmente encontrada na superfície celular de formas amastigotas de *T. cruzi* e, por isso, chamadas de amastinas (TEIXEIRA *et al.*, 1994). Posteriormente, esta família foi caracterizada em *Leishmania* spp. apresentando grande variabilidade, apesar de uma topologia comum, com quatro regiões transmembrana e uma assinatura com 11 aminoácidos conservados (ROCHETTE *et al.*, 2005). Os genes para estas proteínas foram, ainda, encontrados em outros tripanossomatídeos e classificadas em quatro subfamílias (α , β , γ e δ) (JACKSON, 2010). Em *T. cruzi*, seis cópias destes genes, pertencentes às subfamílias β e δ , foram montados em três cromossomos no genoma do clone CL Brener (WEATHERLY *et al.*, 2009).

A fim de caracterizar as diferentes amastinas pertencentes a *T. cruzi*, planejamos a expressão heteróloga destas proteínas em fusão com GFP e a deleção por recombinação homóloga dos genes que estão presentes em cópia única segundo a montagem do genoma do clone CL Brener. Várias outras estratégias comumente usadas em outros organismos para o estudo de função de genes, tem-

se mostrado ineficientes em *T. cruzi*, ou ainda não foram testadas, tornando reduzido o número de ferramentas de manipulação gênica disponíveis. Por isso, este trabalho dedica-se, também, à construção de uma nova ferramenta de nocaute para este organismo, o sistema de recombinação sítio dirigida, mediada por DiCre recombinase, induzida por rapamicina. Este sistema permite a excisão de trechos longos de DNA, especificamente entre dois sítios *loxP* e de forma induzível, sendo possível acompanhar o fenótipo dos parasitos a partir da adição de indutor. Dessa forma, é possível obter de nocautes condicionais, de genes multicópias e de genes essenciais (JULLIEN *et al.*, 2003). O funcionamento deste sistema em *T. cruzi* possibilitará um grande impulso nas estratégias de manipulação gênica neste organismo.

2 REVISÃO BIBLIOGRÁFICA

2.1 DOENÇA DE CHAGAS

A doença de Chagas, cujo agente etiológico é o protozoário *Trypanosoma cruzi*, foi descrita em 1909, quando Carlos Chagas relatou o caso clínico de três crianças portadoras da doença e detalhou o ciclo biológico e a morfologia do parasito (CHAGAS, 1909). A doença é endêmica na América Latina, sendo uma das principais causas de falência cardíaca e com o pior prognóstico quando comparada com outras causas (BOCCHI *et al.*, 2013). No Brasil, no período de 1998 a 2007, foram registrados um total de 9 milhões de óbitos, dos quais 53.930 foram relacionados à doença de Chagas (MARTINS-MELO *et al.*, 2012). Porém, esta patologia não está exclusivamente associada a sua área endêmica. Estima-se que existam 8 a 10 milhões de indivíduos infectados que, devido aos fluxos migratórios, estão espalhados pelo mundo (WHO, 2010, NUNES *et al.*, 2013). Nos Estados Unidos, estima-se que existam 300.000 indivíduos infectados (BERN E MONTGOMERY, 2009), enquanto na Europa este número é de 100.000 (STRASEN *et al.*, 2014).

A transmissão é possível mesmo em regiões não endêmicas ou onde o inseto foi erradicado do ambiente doméstico pois, além da infecção através do vetor, ela se dá de outras formas, como por via oral, transfusão sanguínea, transplantes, via congênita e acidentes de laboratório (NUNES *et al.*, 2013). Apesar das medidas de combate ao inseto terem diminuído consideravelmente o número de casos agudos da enfermidade, esta continua sendo um importante problema de saúde pública. O custo global com a doença está estimado em 7,19 bilhões de dólares por ano, sendo que nas Américas é considerada a doença parasitária mais onerosa (BERN E MONTGOMERY, 2009; NUNES *et al.*, 2013; SCHMUNIS, 2013). Por se tratar de uma epidemia de transmissão vetorial que acomete animais silvestres, domésticos e o homem, a doença de Chagas não pode ser erradicada, mas medidas preventivas

e de tratamentos devem ser aplicadas a fim de diminuir a incidência em humanos (COURA, 2013).

2.2 CICLO BIOLÓGICO DO PARASITO

O parasito *Trypanosoma cruzi* pertence à classe Kinetoplastida e ordem Trypanosomatidae, da qual fazem parte os parasitos *Trypanosoma brucei* e *Leishmania* spp., responsáveis pelas doenças tropicais, também negligenciadas, doença do Sono e Leishmaniose. Em seu complexo ciclo biológico o parasito passa por diversos ambientes e formas biológicas, sendo este ciclo dividido principalmente em duas etapas, uma no hospedeiro mamífero e outra no inseto vetor, um hematófago da subfamília Triatominae, como mostra a figura 1. Ao se alimentar de sangue do mamífero infectado, o inseto ingere parasitos na forma tripomastigota sanguínea que, no intestino anterior do triatomíneo, se diferencia em epimastigota (forma replicativa). Já no intestino posterior do inseto, através do processo de metaciclogênese, o parasito diferencia-se em tripomastigota metacíclico, a forma não-replicativa e infectante para o hospedeiro mamífero, eliminada nas fezes do inseto enquanto o mesmo se alimenta. Quando as fezes contaminadas entram em contato com a pele não íntegra, solução de continuidade ou mucosa do hospedeiro mamífero a infecção pode ocorrer. Nesta forma, o parasito é capaz de invadir uma grande variedade de células. Uma vez dentro da célula, o parasito diferencia-se em forma amastigota que é replicativa e se transforma em tripomastigota, que rompe a célula hospedeira e é liberada no espaço intercelular, podendo atingir a corrente sanguínea. Subsequentemente, o parasito é capaz de infectar novas células, causando a progressão clínica da doença (DE SOUZA, 2002; TYLER E ENGMAN, 2001).

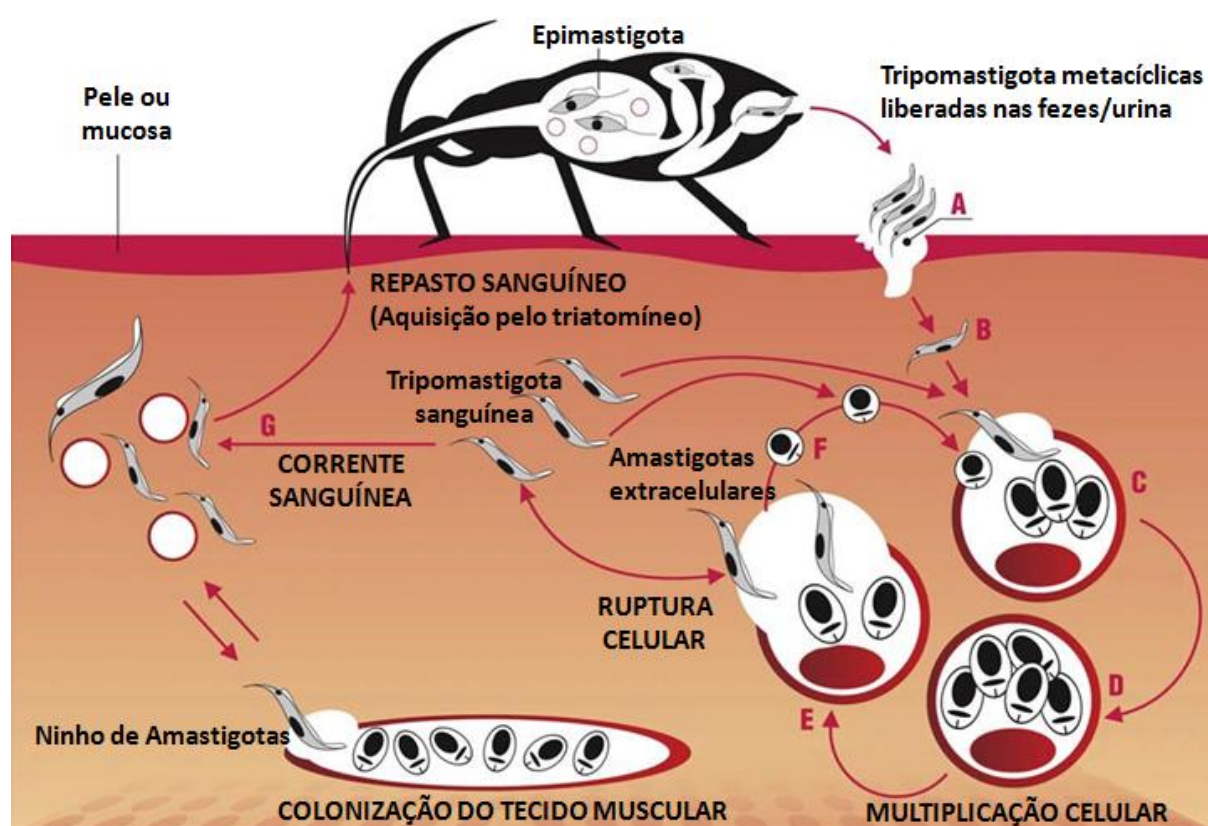


FIGURA 1 – CICLO DE VIDA DO *Trypanosoma cruzi*

FONTE: Adaptado de LIMA *et al.*, 2010

NOTA: Epimastigotas no intestino médio do vetor, um triatomíneo hematófago (acima), diferenciam-se em tripomastigotas metacíclicos que são liberados nas fezes durante o repasto sanguíneo. No contato das fezes com lesão na pele ou com mucosas (A), o parasito na forma infectiva penetra a vizinhança de células hospedeiras (B). Parasitos internalizados permanecem por algumas horas no vacúolo parasitóforo, do qual escapam e se diferenciam na forma replicativa, amastigota, que cresce livre no citoplasma (C). Após alguns ciclos, as amastigotas se diferenciam novamente em tripomastigotas, que rompem a célula infectada e disseminam a infecção (F) ou infectam outro inseto vetor (G). Alternativamente, amastigotas que são liberadas prematuramente de células ou derivam de diferenciação extracelular de tripomastigotas, podem também invadir e manter o ciclo de infecção (E, F)

No momento da lise celular, o parasito pode também ser liberado na forma amastigota, ou ainda se diferenciar em amastigota no ambiente extracelular. Esta forma biológica corresponde a 10% dos parasitos circulantes na corrente sanguínea de camundongos durante a infecção aguda e é capaz de infectar novas células (FERREIRA *et al.*, 2012)

A doença de Chagas pode apresentar diferentes características clínicas e níveis de severidade que dependem, entre outros fatores, das particularidades da

população de parasitos com a qual o paciente se infectou. Entre estas particularidades, inclui-se a variabilidade gênica, que é alta entre as cepas e clones de *T. cruzi*. Baseado nesta variabilidade, estas cepas e clones foram divididos em seis grupos (TcI - TcVI), que podem ser identificados por marcadores imunológicos ou moleculares, dos quais, o grupo TcI corresponde às cepas encontradas em ciclo silvestre, TcII às de ciclo doméstico, enquanto TcIII-VI são grupos compostos por cepas ou clones mistos, como é o caso do clone CL Brener (ZINGALES *et al.*, 2009).

Entre as cepas comumente utilizadas em laboratório, as cepas Brazil Colombiana, Dm, G e Sylvio pertencem ao grupo TcI. No grupo TcII encontramos as cepas Esmeraldo-cl3 e Y. As cepas CL, CL Brener e Esmeraldo estão no grupo TcVI. O perfil gênico e fenotípico é bastante variável entre os grupos, mesmo entre as cepas alocadas no mesmo grupo, como a cepa G e o clone Sylvio X-10, cujas formas amastigotas extracelulares de tecido apresentam reatividades diferentes contra anticorpos anti-amastigotas (VERBISCK *et al.*, 1998). Foi constatado que apesar de haver forte sintonia entre os genomas das diferentes cepas de *T. cruzi*, o conteúdo do elemento retrotransponível L1Tc e do tamanho do genoma foram muito variáveis entre as cepas de diferentes grupos e mesmo entre as três avaliadas dentro do grupo TcI (G, Dm28c e Tc1161) (DE SOUZA *et al.*, 2011). Estas observações justificam o estudo de função de genes em diferentes cepas, já que a expressão gênica, a organização genômica ou o fenótipo de nocautes ou superexpressão de genes pode ser diferente entre as cepas, podendo inclusive explicar a diferença de comportamento entre elas.

Para se adaptar em ambientes tão diversos e diferenciar-se entre uma forma biológica e outra, o parasito expressa proteínas de forma diferencial no decorrer de seu ciclo e, para isso, utiliza mecanismos moleculares peculiares para controle da expressão gênica.

2.3 EXPRESSÃO GÊNICA EM TRIPANOSSOMATÍDEOS

O controle da expressão gênica nos tripanossomatídeos é feito por meio de estratégias um tanto incomuns em outras classes de organismos eucarióticos, como a transcrição policistrônica, processamento do pré-mRNA por *trans-splicing*, editoração de RNA de transcritos mitocondriais e transcrição de genes codificadores de proteína pela RNA polimerase I (TEIXEIRA E DAROCHA, 2003; GOMEZ *et al.*, 2010).

A começar pelo genoma, uma característica particular entre os tripanossomatídeos é a organização, feita em clusters policistrônicos, que contém genes codificadores de dezenas a centenas de proteínas. Estes genes são organizados em conjuntos, mas não codificam proteínas com funções relacionadas entre si, ou seja, não funcionam como os *operons* nos procariotos. Esses conjuntos de genes são transcritos de forma policistrônica e sofrem *trans-splicing* e poliadenilação tornando-se mRNAs maduros monocistrônicos (figura 2), existindo algumas exceções de mRNAs bicistrônicos. Ambas etapas de processamento são guiadas por uma região rica em pirimidinas (PolyPy). Por meio do *trans-splicing* ocorre a adição de *spliced leader*, um pequeno RNA de 39 nucleotídeos “capeado”, à extremidade 5' do transcrito (MARTINEZ-CALVILLO *et al.*, 2010; TEIXEIRA *et al.*, 2012 - Anexo I).

A de organização em clusteres policistrônicos foi evidenciada em vários momentos, incluindo as análises de sequenciamento genômico de tripanossomatídeos (BERRIMAN *et al.*, 2005; EL-SAYED *et al.*, 2005; IVENS *et al.*, 2005; MYLER *et al.*, 1999). O primeiro genoma de *T. cruzi* foi sequenciado a partir do clone CL Brener, quando ficou clara sua característica de clone híbrido. Para a montagem dos alelos, foram usados como referência leituras do clone de Esmeraldo cl3. Assim, um dos alelos do clone híbrido CL Brener é chamado Esmeraldo-Like, e o outro Non-Esmeraldo-Like (ASLETT *et al.*, 2010; DE FREITAS *et al.*, 2006; EL-SAYED *et al.*, 2005). Esta característica híbrida do genoma referência tem dificultado as manipulações da expressão gênica em diferentes laboratórios, já que

genes homólogos podem ter sequências ou regiões intergênicas de sequências diferentes.

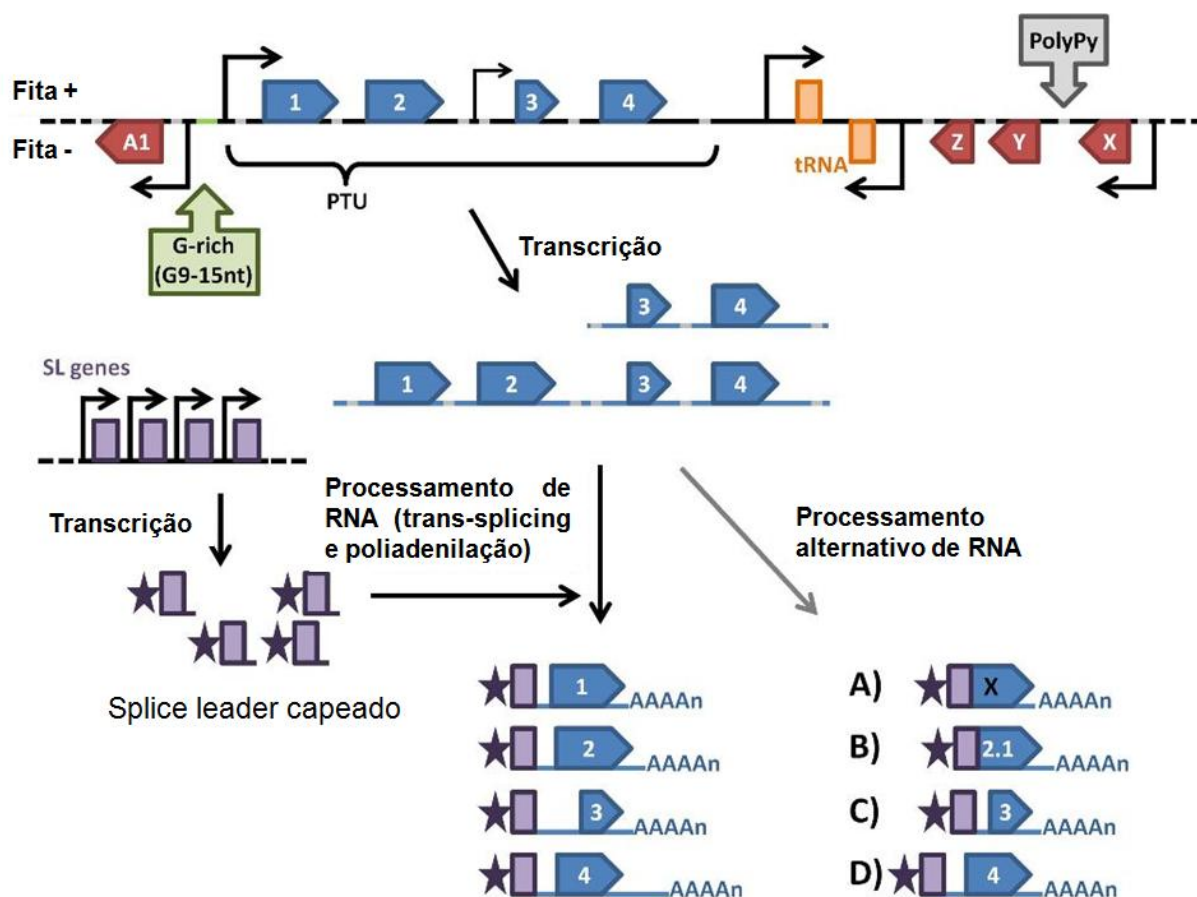


FIGURA 2 - ESQUEMA REPRESENTATIVO DA EXPRESSÃO GÊNICA EM TRIPANOSSOMATÍDEOS

FONTE: adaptado de TEIXEIRA *et al.*, 2012

NOTA: Expressão gênica em tripanossomatídeos. Longos clusters de genes (caixas-seta) não relacionados são organizados em unidades de transcrição policistrônica (PTUs) que são separadas por regiões de mudança de fita transcrita convergentes ou divergentes. Sítios de início de transcrição de RNA polimerase II (TSS) são usualmente localizados a montante do primeiro gene da PTU (MARTINEZ-CALVILLO *et al.*, 2004) ou pode estar localizado internamente (KOLEV *et al.*, 2010). Os RNAs policistrônicos (pré-mRNA) são individualizados em mRNA monocistrônicos após a adição de um RNA *splice leader* capeado por uma reação de trans-splicing acoplada a poliadenilação. Estas reações de processamento são guiadas por uma sequência de polipirimidina presente em cada região intergênica. Os mRNAs maduros são exportados para o citoplasma onde sua tradução e estabilidade podem ser dependentes de elementos *cis*-atuantes presentes em suas regiões não traduzidas (UTR) (ARAUJO *et al.*, 2011). Análises transcriptômicas também mostraram que o pré-mRNA policistrônico pode sofrer processamento alternativo que podem resultar em alterações no códon de início utilizado, consequentemente evitando a tradução (A), ou gerando novas proteínas (B). O processamento alternativo e poliadenilação podem ainda resultar em inclusão ou exclusão de elementos regulatórios presentes na região 5' UTR (C) ou 3' UTR (D), podendo alterar a taxa de expressão de um dado gene (KOLEV *et al.*, 2010; NILSSON *et al.*, 2010; SIEGEL *et al.*, 2010).

A transcrição de genes que codificam proteínas é feita pela RNA polimerase II em eucariotos, enquanto que rRNAs são transcritos pela atividade da enzima RNA polimerase I. Em tripanossomatídeos, alguns genes codificadores de proteínas são transcritos pela RNA pol I (LEE E VAN DER PLOEG, 1997; DANIELS *et al.*, 2010). Evidências sugerem que em *Trypanosoma brucei* a RNA pol I transcreve, além de rRNAs, os genes codificadores de glicoproteínas variantes de superfície (VSG) e de prociclina (GUNZL *et al.*, 2003).

Dentre outros processos peculiares, um bastante interessante é a editoração de RNA, que é conhecida hoje como qualquer modificação no transcrito de DNA mitocondrial, na sequência relacionada ao gene a fim de alterar a região codificadora produzindo um RNA *bona fide* traduzível. O processo é guiado por RNAs pequenos, de aproximadamente 20 nucleotídeos, contendo uma cauda de poliuridina, chamados RNAs guias (gRNA). Estas moléculas pareiam com o RNA pré-editado formando um grampo. A região de não pareamento de bases é alvo de endonucleases. O local de clivagem pode receber ou perder uma uridina, em seguida ocorre uma reação de ligação. Pode ocorrer ainda o processo de edição alternativo, que produz um mRNA diferente do *bona fide* completamente editado, alterando a região codificadora da proteína (HAJDUK E OCHSENREITER, 2010).

A visão geral de formas de controle da expressão gênica, entre outras não exploradas aqui, é crucial no entendimento de como o parasito altera seu “arsenal” de proteínas para suplantar as variadas formas de estresse durante as mudanças de ambientes durante seu ciclo de vida.

2.4 GENES DIFERENCIALMENTE EXPRESSOS DURANTE O CICLO BIOLÓGICO DE *T. CRUZI*

A fim de identificar fatores de virulência, vários pesquisadores têm utilizado diferentes abordagens para identificar genes diferencialmente expressos, tais como,

proteoma, microarranjo, representação da expressão diferencial (RDE), geração de anticorpos específicos contra um estágio, pesquisa em biblioteca de cDNA, RNAseq, entre outros (ANDREWS *et al.*, 1988b; ATWOOD *et al.*, 2005; AVILA *et al.*, 2003; MINNING *et al.*, 2009; NILSSON *et al.*, 2010; SIEGEL *et al.*, 2010; TEIXEIRA *et al.*, 1994).

Avaliando a abundância de transcritos de *T. cruzi* por microarranjo, verificou-se que cerca de 50% dos genes detectados têm sua expressão regulada ao longo do ciclo (MINNING *et al.*, 2009). Em uma abordagem proteômica de larga escala, realizada com extrato proteico do organismo inteiro nos quatro estágios de *T. cruzi*, foram identificados peptídeos mapeados em 2784 proteínas, dentre as quais apenas 834 foram encontradas em todos os estágios (ATWOOD *et al.*, 2005). Sendo assim, o estudo das proteínas diferencialmente expressas pode trazer novos conhecimentos sobre o ciclo biológico e processos de diferenciação, e consequentemente novas alternativas de tratamento ou medidas preventivas contra a doença de Chagas.

Um dos processos de diferenciação mais estudados no *T. cruzi* é a metaciclogênese, quando o parasito passa da forma epimastigota, replicativa e não infectiva, para a forma tripomastigota metacíclica, infectiva para o hospedeiro mamífero. Este processo ocorre no aparelho digestivo do triatomíneo e se dá com grande alteração na composição dos lipídios de membrana e da composição sacarídica da superfície celular, seguida de alterações morfológicas. Diversas proteínas foram identificadas como mais expressas durante a metaciclogênese *in vitro* por meio de um método chamado representação da expressão diferencial (RDE) (AVILA *et al.*, 2003; KRIEGER E GOLDENBERG, 1998) e também por um método mais recente de incorporação de aminoácidos marcados com isótopos estáveis (SILAC) pelo qual foram identificadas 3000 proteínas expressas durante a metaciclogênese (DE GODOY *et al.*, 2012). Como exemplo temos a Metaciclina II e III, localizadas por imunocitoquímica no cinetoplasto e nucléolo, ainda com funções desconhecidas (YAMADA-OGATTA *et al.*, 2004). Posteriormente, foram obtidas algumas informações adicionais para Metaciclina III. Foi confirmada sua localização em pontos do nucléolo e atribuída aos fragmentos N- e C- terminais contendo

aminoácidos básicos (GLUENZ *et al.*, 2007). Foi encontrada, também, uma família de proteínas com domínio de ligação a quitina, um polissacarídeo presente no intestino de triatomíneos, que pode estar associado à adesão do parasito a parede intestinal do inseto (DALLAGIOVANNA *et al.*, 2001).

Algumas proteínas diferencialmente expressas já foram caracterizadas na interação parasito-hospedeiro, especialmente no processo invasivo (ALVES E MORTARA, 2009). Com papel indispensável para este processo, as mucinas são as glicoproteínas expressas em maior quantidade na superfície de *T. cruzi*, com tamanho e conteúdo glicano característico para cada fase do ciclo. Sendo ancoradas por GPI, contêm 50 a 200 aminoácidos, com sequências ricas em serina e treonina, que são sítios de O-glicosilação. Durante a diferenciação ocorre a adição de uma N-acetilglucosamina, que pode ser adicionado de até 5 resíduos de galactose. As extremidades destas porções glicanas podem ser conjugadas a ácido siálico em uma reação catalisada por trans-sialidases (revisado por BUSCAGLIA *et al.*, 2006). O parasito não possui a maquinaria necessária para a síntese de ácido siálico, ao contrário das células hospedeiras, por isso, as trans-sialidades transferem estas moléculas do hospedeiro para o parasito com a finalidade de mimetizar a superfície das células. Estas proteínas são ancoradas a glicofosfatidil inositol (GPI) e atuam mais intensamente em tripomastigota. Elas começam a ser expressas em formas epimastigotas, na fase estacionária, mas sua atividade nesta fase é muito mais baixa que a encontrada no estágio infectivo (revisado por GIORGI E DE LEDERKREMER, 2011).

Outra família de proteínas preferencialmente expressa no estágio infectivo é representada pelas proteínas associadas a mucinas (MASPs). Os genes desta família correspondem a 6% do genoma de *T. cruzi* e estão organizados próximos a genes de mucinas, trans-sialidases e elementos de transposição. As proteínas preditas desta família têm alta conservação nas extremidades N- e C-terminal e grande variação em tamanho e sequência na região central (BARTHOLOMEU *et al.*, 2009). Existem evidências de que o parasito varie o gene de MASP que está sendo expresso durante a infecção e dentre os indivíduos da população de parasito. Esta

alternância de expressão a grande variabilidade dos genes desta família sugerem que a função destas proteínas pode estar relacionada com a evasão do parasito do sistema imunológico (DOS SANTOS *et al.*, 2012).

Ainda tratando-se da relação parasito-hospedeiro, uma proteína bastante conhecida por atuar neste contato é a gp63. Ela está presente no parasito em dois tamanhos diferentes, sendo que em tripomastigota prevalece a de maior tamanho (78 kDa), enquanto a menor (75 kDa) é mais expressa em epimastigotas, e diminui a medida que a cultura envelhece. Esta diferença de tamanho pode estar associada a glicosilação, que é mais necessária no estágio infectivo, protegendo o parasito contra lise pelo sistema complemento. Esta proteína apresentou atividade de metaloprotease e a incubação do parasito com anticorpos anti-gp63 levou ao bloqueio parcial da invasão de células hospedeiras por tripomastigotas, o que indica que gp63 interage com o hospedeiro facilitando a infecção (CUEVAS *et al.*, 2003; KULKARNI *et al.*, 2009).

Outra glicoproteína de membrana, agora com função estrutural/motora, é chamada gp72. A proteína é mais expressa em epimastigota, e foi mostrado por nocaute do gene que esta é responsável pela adesão do flagelo ao corpo do parasito (COOPER *et al.*, 1993). Outra função provável desta proteína está no desencadeamento da metaciclogênese, já que o bloqueio desta por anticorpo inibe a diferenciação (SHER E SNARY, 1982). Recentemente, foi descrito o fosfossacarídeo ligado a esta glicoproteína, um epítipo altamente imunogênico e de alta complexidade, cuja estrutura pode colaborar nas funções atribuídas a gp72 (ALLEN *et al.*, 2013).

Entre as proteínas de superfície que foram descritas preferencialmente expressas em amastigota, podemos destacar as ASP-1 e ASP-2 (SANTOS *et al.*, 1997 LOW E TARLETON, 1997) que são membros da família das *trans*-sialidases, SA-85 (KAHN *et al.*, 1990 KAHN *et al.*, 1991), SSP4 (ANDREWS *et al.*, 1988a; OLIVAS-RUBIO *et al.*, 2009) e amastinas (TEIXEIRA *et al.*, 1994). A maior parte dessas proteínas são glicoproteínas ancoradas por GPI à membrana e foram identificadas utilizando anticorpos gerados contra formas amastigotas. No caso de

SSP4, os anticorpos gerados reconhecem epítomos de carboidratos desta glicoproteína (ANDREWS *et al.*, 1988a). Em outro estudo, Teixeira *et al.* (1994), utilizando uma bibliotecas de cDNA, identificaram genes que codificam para proteínas de superfície preferencialmente expressas no estágio intracelular do parasito, as quais foram denominadas de amastinas. O papel das proteínas codificadas por estes genes ainda é desconhecido. Sendo uma proteína exclusiva de tripanossomatídeos, diferencialmente expressa na forma replicativa presente no hospedeiro mamífero, e localizada na superfície do parasito, pode tornar-se um alvo de drogas ou antígeno para produção de vacinas. Por isso, nosso grupo de pesquisa está interessado na caracterização funcional destes genes.

2.5 AMASTINAS

As amastinas foram inicialmente caracterizadas como proteínas de superfície preferencialmente expressas em amastigotas de *T. cruzi*, codificadas por genes multicópias (TEIXEIRA *et al.*, 1994). Posteriormente, sequências de amastinas foram encontradas em outros tripanossomatídeos como *T. brucei* e *Leishmania* spp., que foram comparadas entre si para caracterização estrutural. (ROCHETTE *et al.*, 2005). De forma geral, a estrutura secundária das amastinas apresenta quatro regiões transmembrana, sendo que a primeira região corresponde a um potencial peptídeo sinal, duas regiões extracelulares e as extremidades voltadas para o citoplasma (figura 3). Analisando amastinas de *Leishmania* spp., *Trypanosoma cruzi* e *T. brucei*, foram encontrados aminoácidos com alto índice de conservação. Em *Leishmania major*, a localização de amastina na superfície celular foi mostrada por expressão desta em fusão com proteína verde fluorescente (GFP) em cultura amastigota axênica (ROCHETTE *et al.*, 2005). Em *T. cruzi*, a mesma localização foi encontrada (CRUZ *et al.*, 2012; TEIXEIRA *et al.*, 1994).

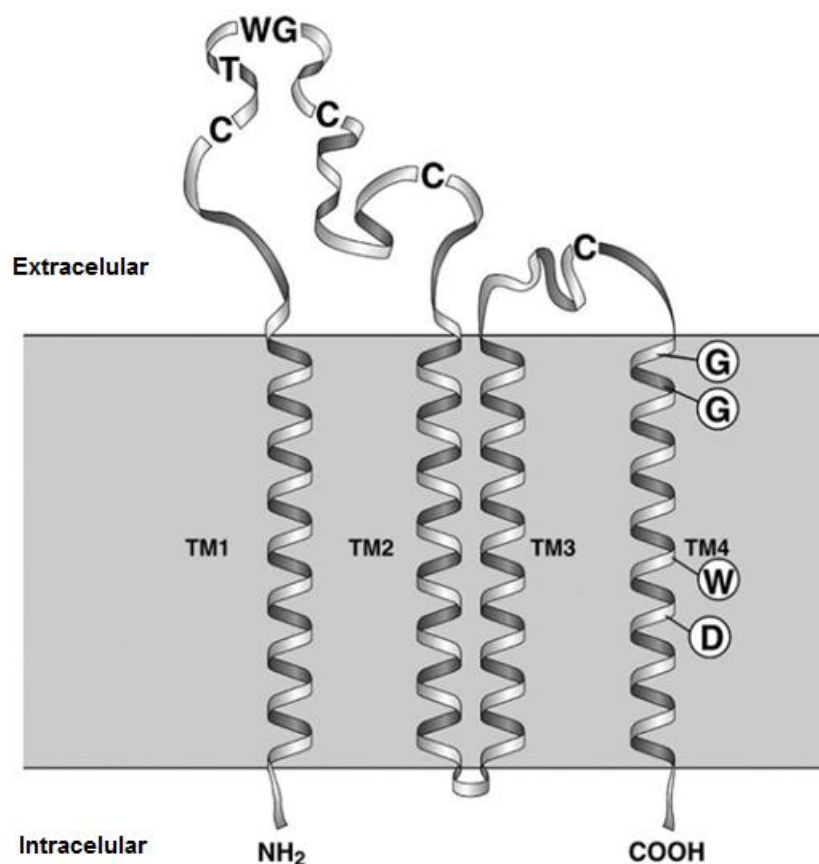


FIGURA 3 – TOPOLOGIA PREDITA DAS PROTEÍNAS DE SUPERFÍCIE AMASTINAS

FONTE: Adaptado de ROCHETTE *et al.*, 2005

NOTA: A maioria das amastinas contém quatro regiões transmembrana (TM) com N-terminal curto e cauda C-terminal, ambos voltados para o citosol. A quarta região transmembrana é melhor estruturada com quatro resíduos altamente conservados. Homólogos do gene amastina contém dois domínios extracelulares. O primeiro domínio delimitado entre TM1 e TM2 contém 55-60 aminoácidos de comprimento e contém vários aminoácidos universalmente conservados. A região delimitada entre as duas cisteínas é altamente conservada entre todos os genes homólogos de amastina em *Leishmania* spp. e *Trypanosoma* spp. e é exclusivo para esta família de proteínas (assinatura da amastina). O segundo domínio extracelular entre TM3 e TM4 é menor, menos conservado e de tamanho mais variável. Letras representam aminoácidos conservados.

Foram encontrados genes para amastinas em diversos tripanossomatídeos que tiveram seu genoma sequenciado (JACKSON, 2010). Por meio de uma análise filogenética das sequências de amastinas de *Leishmania major*, *Leishmania infantum*, *Leishmania braziliensis*, *Leptomonas seymouri*, *Angomonas* sp., *Trypanosoma cruzi* e *Trypanosoma brucei* foi proposta uma classificação, na qual os genes foram divididos em 4 subfamílias: α , β , γ e δ . Além da proximidade

filogenética, alguns atributos comuns caracterizam estas subfamílias. As α -amastinas aparecem como um par de genes distintos em *tandem*. As β - e γ -amastinas também se apresentam como pares de genes distintos, porém podem estar presentes em múltiplas cópias. Segundo Jackson (2010), estas duas subfamílias se diferem, pois β -amastinas estão conservadas em vários tripanossomatídeos e não somente em *Leishmania* sp., como as γ -amastinas. As δ -amastinas estão presentes em vários locus, o mesmo gene em *tandem* alternado com o gene de tuzina, sendo que há uma expansão no número de cópias desses genes no gênero *Leishmania* (JACKSON, 2010; ROCHETTE *et al.*, 2005).

Análises filogenéticas comparando genes de amastinas de *Leishmania* (L.) *amazonensis* com 181 genes de outras espécies de *Leishmania* sp. mostraram que existem amastinas exclusivas para as espécies de *L. (L.) major*, *L. (V.) braziliensis* e *L. (L.) infantum*, e três grupos em que as amastinas de *L. (L.) amazonensis* e *L. (L.) mexicana* compõem ramos distintos. Esta divisão sugere que algumas amastinas surgiram devido à pressão do ambiente ou por especiação do patógeno, e pode estar relacionada às diferenças de manifestação clínica da doença causada pelas espécies avaliadas (REAL *et al.*, 2013).

Com base na anotação do genoma de CL Brener disponíveis no banco de dados TriTrypDB (2014), e informações retiradas de Jackson (2010), em *Trypanosoma cruzi* existem 12 genes (6 de cada alelo) anotados como amastinas e localizadas em três cromossomos diferentes, pertencentes a duas subfamílias, β (duas cópias) e δ (quatro cópias) como mostrado na figura 4. As δ -amastinas são codificadas por genes presentes nos cromossomos 26 e 34, com uma e três cópias respectivamente, sendo acompanhadas por sequências codificadoras de tuzina. Já as β -amastinas são codificadas por duas cópias, não acompanhadas pelo gene de tuzina, presentes no cromossomo 32, arranjadas em sequência com uma grande região intergênica entre elas (TRITRYP DB, 2014). Estas subfamílias serão parcialmente caracterizadas neste estudo. Cabe ressaltar que recentemente nosso grupo de pesquisa identificou uma nova subfamília em *T. cruzi*, as α -amastinas, que já foram parcialmente caracterizadas por Lemos (2013).

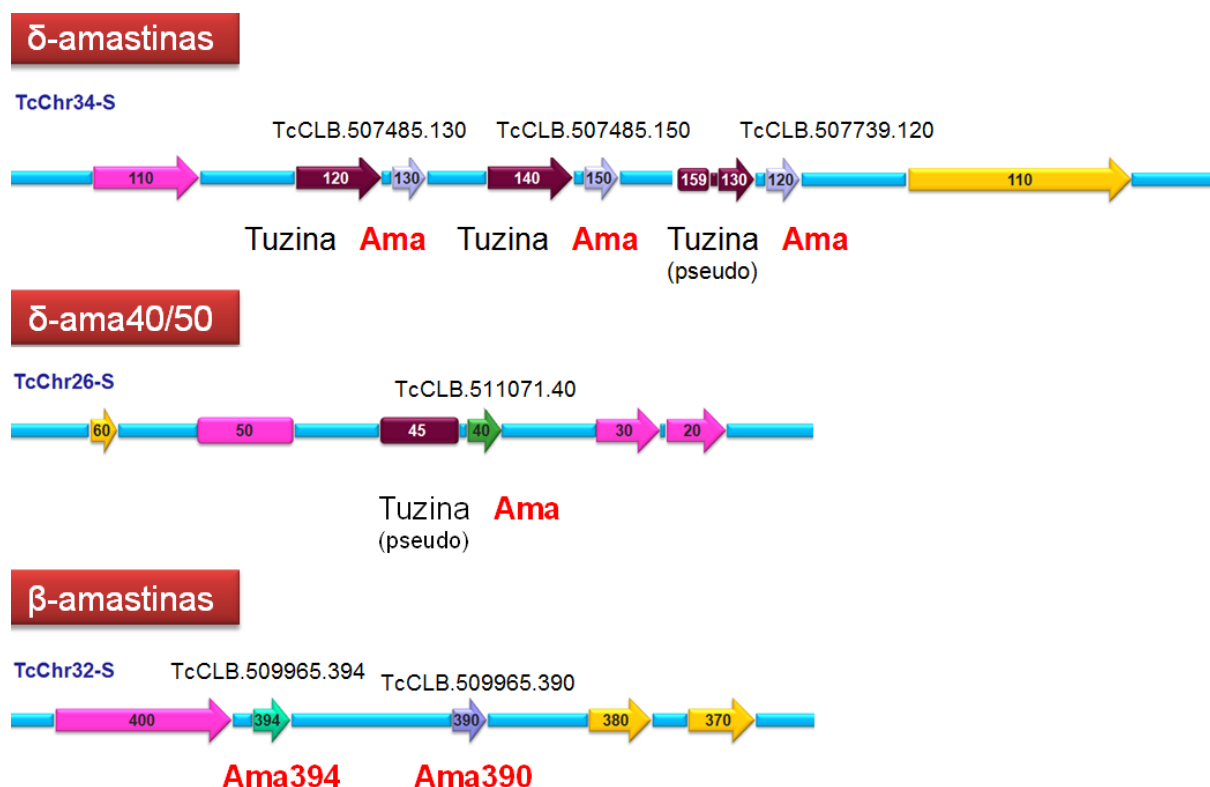


FIGURA 4 – ORGANIZAÇÃO GENÔMICA DAS ISOFORMAS DAS AMASTINAS DE *T. cruzi*

FONTE: Adaptado de TRITRYPDB, 2014 e JACKSON, 2010.

NOTA: As seis amastinas encontradas em *Trypanosoma cruzi* são pertencentes a duas subfamílias, β e δ . As δ -amastinas são codificadas por genes presentes nos cromossomos 26 (δ -Ama40) e 34, com uma e três cópias respectivamente, sendo acompanhadas por sequências codificadoras de tuzina. As β -amastinas são codificadas por duas cópias do gene presentes no cromossomo 32 (JACKSON, 2010; TRITRYPDB, 2011). As setas indicam regiões codificadoras de proteínas. As setas correspondentes a amastinas apresentam acima seu código identificador. As caixas indicam pseudogenes.

Dentre as amastinas de *T. cruzi*, até o momento, a melhor estudada é a subfamília das δ -amastinas, que são as amastinas descritas por TEIXEIRA *et al.*, 1994). Algumas evidências sugerem que elementos de regulação pós-transcricional localizados na região 3'UTR sejam responsáveis pela expressão diferencial (COUGHLIN *et al.*, 2000; MCNICOLL *et al.*, 2005). Um elemento de 43 nucleotídeos rico em U foi encontrado na região 3'UTR de um grande número de mRNA mais abundantes em amastigota de *T. cruzi*. Este elemento é alvo de ligação para a proteína TcUBP1, que liga a região 3'UTR do mRNA de amastina (DE GAUDENZI *et*

al., 2003; LI *et al.*, 2012). Em um estudo recente verificou-se que a depleção de proteínas alba20 em formas amastigotas de *Leishmania infantum* leva a uma diminuição da estabilidade dos transcritos de δ -amastinas.

Uma vez que a forma amastigota é a responsável pela replicação do parasito no paciente chagásico, as δ -amastinas, com maior expressão nesta etapa, têm sido avaliadas como bons antígenos na imunização de animais e proteção contra infecção parasitária. Alguns estudos em *Leishmania* SP. e *T. cruzi* já obtiveram sucesso nesta aplicação, sugerindo que essas proteínas podem estar relacionadas com a infectividade do parasito. Após serem imunizados com vetores contendo genes codificadores de diversas proteínas de *L. major*, os camundongos apresentaram maior resistência a infecção do mesmo parasito quando imunizados com genes de amastina comparando-se com outras proteínas (STOBER *et al.*, 2006). A expressão de amastina em fusão com as proteínas V22 do vírus da herpes e EGFP melhoraram a imunização de camundongos contra *L. major* (BOLHASSANI *et al.*, 2011). Para *T. cruzi* foram encontradas proteínas de superfície capazes de imunizar camundongos, como as trans-sialidases. Esta imunização é melhorada com co-administração de interleucina IL-15 (EICKHOFF *et al.*, 2011).

Recentemente, análises de expressão diferencial apontaram as amastinas como fatores de virulência de *T. cruzi*. No seguinte estudo, foram avaliadas formas amastigotas extracelulares (AE), que são diferenciadas em meio ácido a partir de tripomastigota de obtido de cultura (TCT). A capacidade de invasão de células em cultura é maior para a forma AE proveniente de Cepa G quando comparada à do clone CL Brener. Por sua vez, a expressão de mRNA de δ -amastina é menor para a cepa G do que para o clone CL Brener. A superexpressão de δ -amastinas na Cepa G diminuiu a capacidade invasiva se da forma AE, tornando seu perfil mais próximo ao do clone CL Brener, o que sugere que a expressão de δ -amastina dificulta o processo invasivo de formas AEs em cultura de tecido. Além disso, foi proposto neste trabalho que a expressão das δ -amastina em fusão à GFP em epimastigotas de cepa G contribuiu para a sobrevivência do parasito, por acelerar a metaciclôgenese e na conversão de amastigotas para tripomastigota em cultura de

tecido (TCT), além de tornar a parasitemia tecidual precoce no modelo animal (CRUZ *et al.*, 2012).

Este controle diferenciado da expressão das subfamílias de amastinas, sua estrutura e localização, indicam que esta classe de proteínas pode ter uma função importante, mas não essencial, na biologia do parasito. Desta forma, conhecer a função desta família de proteínas é interessante, já que pode tratar-se de um alvo de drogas ou antígeno para produção de vacinas. Contudo, o estudo de função de genes, principalmente de famílias multigênicas, em *T. cruzi* é, em grande parte, limitado devido ao reduzido número de ferramentas de manipulação gênica disponíveis, fato que está intimamente relacionado às diversas características peculiares que estes parasitos apresentam em sua biologia molecular.

2.6 FORMAS DE MANIPULAÇÃO DA EXPRESSÃO GÊNICA EM TRIPANOSSOMATÍDEOS

Devido a peculiaridades em sua biologia molecular, tais como a presença de diversos genes em múltiplas cópias e ausência de funcionalidade da maquinaria de RNAi, a manipulação gênica destes organismos demanda cuidados especiais, que muitas vezes impede o uso de estratégias de perturbação da expressão gênica comumente usadas em outros organismos. Em *T. cruzi*, a expressão heteróloga de proteínas corresponde à maior parcela entre as poucas ferramentas bem sucedidas. Em relação ao bloqueio da expressão, a única opção já utilizada é a geração de nocautes por recombinação homóloga, que é laboriosa e limitada a conjuntos de genes cópia simples e não essenciais. Para outros tripanossomatídeos outras ferramentas estão disponíveis, como o bloqueio da expressão por RNAi e a expressão ou nocaute induzidos.

O primeiro plasmídio bem sucedido em expressar um gene exógeno em Trypanosomatidae foi o pAL5, desenvolvido contendo *spliced leader* na porção 5' e sinal para adição de cauda poli-A na porção 3' do gene de cloranfenicol acetil transferase (CAT). Nesta situação, foi encontrada atividade transiente de CAT em

Leptomonas seymori (BELLOFATTO E CROSS, 1989). A primeira tentativa de encontrar uma região promotora para expressão gênica foi realizada para *Leishmania* sp.. A região intergênica da α -tubulina foi inserida a montante do gene de CAT no vetor pBlueScript, mantendo o *spliced leader* e o sinal para adição de cauda poli-A. Este plasmídeo foi chamado pALT1-1 (LABAN E WIRTH, 1989).

Mais tarde, foi criado o primeiro vetor não integrativo para *T. cruzi*, carregando um gene repórter e um gene de resistência à neomicina, chamado pTEX (KELLY *et al.*, 1992). Com a adição de um fragmento de 800pb correspondente a um promotor de RNA ribossômico antes do sítio de policlonagem foi construído o pRIBOTEX, para o qual o período de seleção de resistentes foi encurtado em relação ao vetor anterior (MARTINEZ-CALVILLO *et al.*, 1997). A partir desta construção, o nível de expressão da proteína recombinante foi melhorado, com a inserção de parte da região 5'UTR da proteína ribossômica TcP2 β (HX1) logo após o promotor de rRNA. A construção gerada, o vetor pTRES, é muito utilizada atualmente para expressar genes repórteres ou proteínas fusionadas a "tags" (VAZQUEZ E LEVIN, 1999). Como uma ferramenta alternativa para a expressão estável de proteínas recombinantes no parasito, foi construído um vetor para integração no genoma do parasito no locus de β -tubulina, pROCKGFPNeo, que carrega os genes de proteína verde fluorescente (GFP) e de resistência a neomicina. O gene de GFP pode ser substituído ou fusionado a qualquer gene de interesse (DAROCHA *et al.*, 2004b).

Um recurso que seria muito aplicável ao estudo de função de genes é a expressão induzida, tanto para geração de nocautes induzíveis como para expressão de proteínas tóxicas ao parasito. Em *T. brucei*, o sistema de indução da expressão por tetraciclina foi padronizado com a expressão da T7RNA polimerase e do repressor de tetraciclina (TetR) em uma linhagem resistente a higromicina e neomicina. Os plasmídios utilizados para gerar esta linhagem são pLEW13 e pLEW29 respectivamente, cuja numeração corresponde ao cromossomo ao qual eles se integram (WIRTZ *et al.*, 1999).

Na tentativa de construir um sistema similar, o vetor pLEW13 foi modificado para conter também o cassete de expressão do TetR, além do promotor de rRNA de

T. cruzi. Sabendo que a região de recombinação (gene de β -tubulina) apresenta 90% de similaridade entre as duas espécies, *T. brucei* e *T. cruzi*, o plasmídeo foi transfectado e a população resistente contendo esta construção foi chamada CL Brener13. A atividade de luciferase desta população passou de 23 para 574 após a adição de tetraciclina. O aumento é considerável, porém não há relato de uso do sistema em trabalhos posteriores, por ter sido considerada baixa a eficiência de regulação, e os possíveis efeitos da inserção do vetor neste locus não foram testados (DAROCHA *et al.*, 2004a).

Em um segundo esforço para construção de um sistema de expressão induzido mais eficiente foi criado o pTcINDEX. Também foi iniciado a partir do pLEW13, porém a região de integração foi alterada para a região espaçadora não transcrita de RNA ribossômico pois esta era a única conhecida por ser transcricionalmente inativa (TAYLOR E KELLY, 2006). O sistema mostrou ser induzível, porém foi relatado vazamento de expressão detectável por *western blot* para a população não induzida (LAVERRIERE *et al.*, 2012; PIACENZA *et al.*, 2007). A fim de buscar novas formas de regulação da expressão, que poderiam ser utilizadas como alternativas para diminuir o vazamento, algumas opções vêm sendo estudadas. Por exemplo, foi mostrado recentemente que uma ribozima sensível à tetraciclina pode regular a expressão de um gene repórter (RLUC) em *T. cruzi* (CAMBRI, 2011).

Em relação ao bloqueio da expressão gênica, a única forma usada até hoje em *T. cruzi* foi a deleção por recombinação homóloga, que ainda é bastante complicada devido à presença de vários genes organizados em famílias multigênicas em *tandem* no genoma ou dispersas em cromossomos distintos, além de regiões intergênicas, comumente usadas para recombinação, com sequências repetitivas (ELIAS *et al.*, 2009). Foi conseguido um avanço através do sistema Gateway® (Invitrogen), que torna mais rápida e prática a construção de cassetes para nocaute (XU *et al.*, 2009), mas a estratégia de deleção em si continua com os mesmos obstáculos. Em outros organismos, uma opção para esta dificuldade seria o uso do RNA de interferência (RNAi) no bloqueio da expressão gênica. Porém, *T.*

cruzi é deficiente de alguns componentes da maquinaria desta via de interferência na expressão gênica (DAROCHA *et al.*, 2004a; BATISTA E MARQUES, 2011). Estas dificuldades impulsionam a procura por novos sistemas para deleção ou regulação da expressão gênica. Recentemente, o sistema CRE//*loxP*, uma ferramenta molecular bastante utilizada em alguns organismos, vem sendo testada em *T. brucei* com sucesso (BARRETT *et al.*, 2004; KIM *et al.*, 2013; SCAHILL *et al.*, 2008).

2.7 O SISTEMA CRE-LOXP E DICRE-LOXP

Nesse sistema, a recombinase CRE, uma topoisomerase tipo I codificada pelo bacteriófago P1, catalisa recombinação sítio específica entre dois sítios *loxP* (34 pb) sem necessitar de nenhum cofator. O sítio de recombinação *loxP* corresponde a um elemento de 34 pares de base, que possui duas repetições de 13 nucleotídeos flanqueando uma sequência espaçadora de oito nucleotídeos que confere direcionalidade ao rearranjo. A orientação dos sítios *loxP* afetam a recombinação, podendo levar a deleções e inversões como mostrado na figura 5. A partir do trabalho pioneiro de Brian Sauer e colaboradores, transferindo a tecnologia do sistema de recombinação procariótico para leveduras (SAUER, 1987), vários outros trabalhos têm transferido esse sistema para diferentes tipos celulares incluindo células de mamíferos, *T. brucei*, entre outros (SAUER E HENDERSON, 1988; BARRETT *et al.*, 2004; SCAHILL *et al.*, 2008). O sistema CRE//*loxP* é a base para várias abordagens de modificações em DNA genômico e rearranjos gênicos em camundongo, incluindo nocautes gênicos condicionais (GAVERIAUX-RUFF E KIEFFER, 2007; JULLIEN *et al.*, 2003; LE E SAUER, 2001; MILLS E BRADLEY, 2001; WILSON E KOLA, 2001).

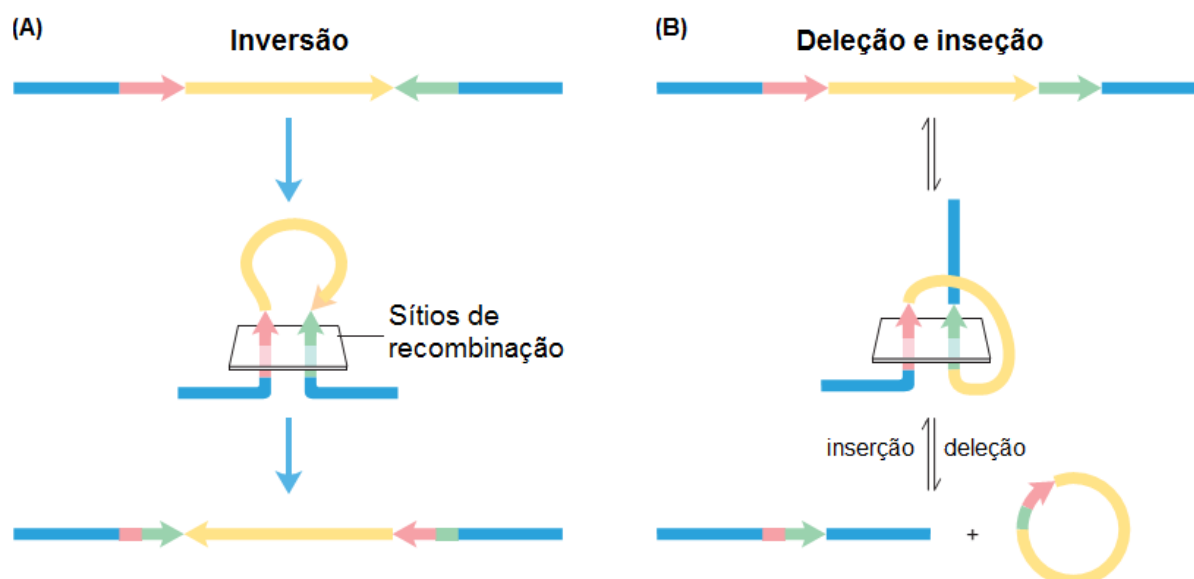


FIGURA 5 – EFEITO DE RECOMBINAÇÃO SÍTIO-ESPECÍFICA

FONTE: Adaptado de NELSON E COX, 2005

NOTA: O resultado da recombinação sítio específica depende da localização e orientação do sítio de recombinação (setas verdes ou cor-de-rosa) em uma molécula de DNA dupla fita. (A) Recombinação com sítios em orientações opostas na mesma molécula de DNA. O resultado é a inversão. (B) Recombinação com sítios na mesma orientação, ou na mesma molécula de DNA, ocasionando deleção ou em duas moléculas de DNA, ocasionando inserção.

É importante destacar a versatilidade deste sistema. Em 2007, foi publicado um trabalho elegante no qual os pesquisadores realizaram evolução *in vitro* da recombinase CRE, tornando-a capaz de reconhecer sequências LTR do HIV e recombinar, ao invés de sequências *loxP*. Desta forma a recombinase resultante funciona como tesoura molecular, levando a excisão do HIV do genoma (SARKAR *et al.*, 2007). A recombinase CRE pode ser utilizada ainda com o objetivo de se estudar a dinâmica de síntese de proteínas. Nesta abordagem, foi criado um sistema de marcação permanente de proteínas recém-sintetizadas após indução da recombinase CRE regulada por ação hormonal. Um cassete foi introduzido no *locus* endógeno do gene de interesse. Proteínas produzidas anteriormente a indução da expressão da recombinase CRE eram traduzidas com uma tag HA. Após a indução, a recombinase removeu o trecho de DNA flanqueado por *loxP*, fazendo com que a região codificadora da proteína fosse fusionada a tag T7. Este sistema foi utilizado no estudo de histonas, a fim de comparar a mobilidade de proteínas previamente

sintetizadas e de proteínas recém-sintetizadas na estrutura da cromatina (VERZIJLBERGEN *et al.*, 2009).

Em outra abordagem, foi criado um nocaute condicional em ratos por meio de uma estratégia sofisticada. Foram inseridos cassetes contendo sítios *loxP* flanqueando o gene *Grin1* utilizando nucleases dedos de zinco (ZFNs). Estas enzimas clivam o DNA em locus específico aumentando a eficiência de recombinação, que no caso deste trabalho gerou embriões contendo os genes de interesse flanqueados por *loxP*. O cassete de expressão da recombinase CRE foi inserido no genoma sob o controle do promotor *Th*, que está ativo somente em alguns tecidos, como a glândula adrenal. Dessa forma, o bloqueio da expressão do gene *Grin1* foi condicionado à atividade do promotor, sendo observado na glândula adrenal e não em outros tecidos, como o cerebral (BROWN *et al.*, 2013).

Em tripanossomatídeos, o sistema já foi utilizado em *T. brucei* com a expressão da recombinase CRE regulada por tetraciclina. Neste trabalho os autores testaram a recombinação de uma sequência terminadora transcricional flanqueada por sítios *loxP*, a montante de um gene repórter. A deleção eficiente da sequência terminadora permitiu a expressão de GFP. Porém foi observada atividade tóxica da recombinase CRE mesmo na ausência de tetraciclina em formas procíclicas de *T. brucei* (BARRET *et al.*, 2004). Em outro trabalho, a fim de diminuir o efeito tóxico da recombinase, o sistema foi utilizado tanto em transfecção transiente quanto estável em *T. brucei* na forma sanguínea, na qual o vetor apresentava uma expressão 10 vezes mais baixa. A recombinação foi eficiente para a retirada de marcas de seleção (genes de resistência a drogas) (SCAHILL *et al.*, 2008).

Em um trabalho mais recente, a recombinase CRE foi utilizada em *T. brucei* para retirada de marcas de seleção e nocaute condicional ao mesmo tempo, em uma estratégia interessante para nocaute de genes essenciais. Foram construídos dois cassetes, um para o nocaute clássico, com o gene de resistência a puomicina flanqueado por *loxP* e outro, com as mesmas regiões de recombinação, para deleção do segundo alelo. Neste, a região flanqueada por *loxP* continha o gene de resistência a Higromicina e a ORF do gene nocauteado marcado com epítipo Ty1.

Ambos os genes de resistência foram expressos em fusão com HSVTK (Herpes simplex virus thymidine kinase), que confere sensibilidade a ganciclovir e com epítipo de Ty1. Os dois cassetes foram transfectados no parasito e selecionados com os antibióticos apropriados. Após a expressão de recombinase CRE foi obtida uma população sensível a puromicina e a higromicina e resistente a ganciclovir. Foi mostrado por *western blot* que a expressão das proteínas marcadas com epítipo de Ty1 foi bloqueada (KIM *et al.*, 2013).

Há uma preocupação com a expressão da recombinase CRE, que se deve a toxicidade da proteína, relatada em diversos trabalhos (BARRETT *et al.*, 2004; HEIDMANN E LEHNER, 2001; SCAHILL *et al.*, 2008). Em um trabalho mais atual, foi mostrado que a expressão da recombinase CRE sem a presença de sítios *loxP* gera danos ao DNA, instabilidade do genoma e tetraploidia em queratinócitos de pele de camundongo expressando este gene, além de alteração no ciclo celular de células de carcinoma colorretal, da linhagem HCT116 (JANBANDHU *et al.*, 2014). Por isso, é desejável obter um controle mais fino da expressão desta proteína. Com este intuito, foi criado o sistema DiCre, no qual a recombinase é expressa em duas porções que se aproximam somente na presença de rapamicina, tornando-se, então ativa (figura 6). Desta forma, a toxicidade da enzima está limitada ao período de indução (JULLIEN *et al.*, 2003). Assim, este sistema apresenta a versatilidade do sistema CRE/*loxP*, possibilitando a excisão ou inserção de grandes trechos de DNA, e além de oferecer menor toxicidade à célula, tem a vantagem de ser induzível, o que possibilita a geração de nocautes condicionais. Por estas razões, o sistema DiCre/*loxP* foi escolhido para ser testado neste trabalho.

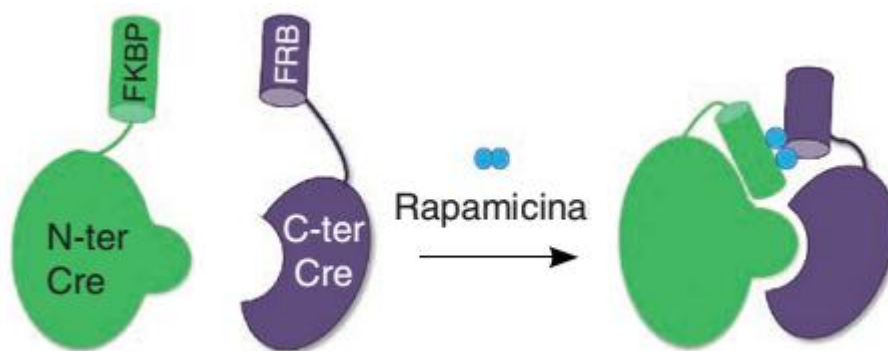


FIGURA 6 – ESQUEMA REPRESENTATIVO DO SISTEMA DICRE

FONTE: Adaptado de ANDENMATTEN *et al.*, 2013

NOTA: A proteína é dividida em dois fragmentos sem atividade enzimática que são fusionados aos domínios FKBP e FRB que dimerizam na presença de um ligante. Esta dimerização leva a aproximação dos dois fragmentos da recombinase Cre, que retoma sua atividade enzimática.

Um controle mais fino da atividade da recombinase CRE pode ser obtido através de sistemas regulados, como foi mostrado em camundongos com o sistema CRE dimerizável (DiCre). Nesse sistema, o gene da recombinase CRE foi dividido em duas partes, sendo que uma parte que codifica a proteína até o aminoácido 59 (Cre59), e a outra codificando a partir do aminoácido 60 até o último (Cre60). A Cre59 foi fusionada com o domínio FRB, enquanto a Cre60 com o domínio FKBP12. Adicionalmente, estas proteínas truncadas também foram fusionadas a um peptídeo de localização nuclear permitindo endereçamento da proteína de fusão para o núcleo. As duas partes da recombinase CRE separadas são mantidas próximas na presença de rapamicina, tornando a recombinase ativa (JULLIEN *et al.*, 2003). Este sistema foi funcional quando a construção foi inserida para expressão de CRE dimerizável em camundongos no locus Rosa26. Foi também bastante eficiente na presença de rapamicina, não apresentando atividade na ausência deste indutor, constituindo um modo interessante de promover deleção condicional de um trecho de DNA (JULLIEN *et al.*, 2007).

O sistema DiCre foi utilizado também em *Toxoplasma gondii* com objetivo de bloquear totalmente a expressão de genes e verificar se o fenótipo seria o mesmo encontrado anteriormente, quando estes genes tiveram sua expressão reduzida por outra metodologia (sistema de transativação induzido por tetraciclina). A estratégia

usada para o sistema DiCre consistiu na inserção de um cassete com sítios *loxP* flanqueando o gene de interesse. Em uma das estratégias usadas no trabalho, após a recombinação induzida por rapamicina, o gene era nocauteado e o parasito passava a expressar a proteína amarelo fluorescente. Após 24 horas de indução com 50 nM de rapamicina foi observada 96% de fluorescência (ANDENMATTEN *et al.*, 2013).

Este sistema de regulação permite a geração de nocautes condicionais mesmo na ausência de controle transcricional dos domínios de recombinase CRE. Esta nova abordagem de deleção condicional (regulada ou controlada) permite a deleção de genes essenciais em população de células de forma mais sincronizada que a geração de nocautes convencionais, possibilitando assim a análise de fenótipos de forma mais acurada. Uma vez funcional em *T. cruzi*, este sistema ampliaria as possibilidades de estratégias de deleção desde genes cópia simples até a remoção de clusters gênicos inteiros de seu genoma. O uso deste sistema para retirada de marcas de resistência também poderia ser amplamente aplicado, já que por existirem poucas opções para *T. cruzi*, o mesmo gene de resistência poderia ser reutilizado. Sendo assim muitos progressos seriam possíveis nas estratégias para obtenção de nocautes por recombinação homóloga de genes com múltiplas cópias, ou na complementação de nocautes duplo-resistentes. Sendo assim, esta ferramenta de manipulação gênica poderá permitir o estudo funcional de vários genes, especialmente para pequenas famílias multigênicas, como é o caso das amastinas.

3 JUSTIFICATIVA

Proteínas diferencialmente expressas em formas infectivas podem apresentar funções importantes na virulência, e na diferenciação e/ou adaptação às novas condições, sendo, portanto, um interessante alvo de estudos, já que o bloqueio ou interferência na função de algumas destas proteínas pode culminar no insucesso da infecção, permitindo o controle da doença. Frente ao cenário atual de reduzidas possibilidades de tratamento para doença de Chagas e com efeitos colaterais severos, é importante a descoberta de alvos que possibilitem tratamentos alternativos e/ou desenvolvimento de vacinas. Para este fim, proteínas chaves na infecção e que sejam distintas de proteínas do hospedeiro, são as mais interessantes. Neste sentido, a família de proteínas amastinas parece preencher todos esses requisitos. Tratam-se de proteínas exclusivas de tripanossomatídeos, que quando utilizadas em protocolos de vacinação geram proteção. Por estarem na superfície celular, pode facilitar o acesso de drogas e/ou acesso a anticorpos gerados em protocolos de imunização.

Com base em dados de genoma disponíveis e dados previamente publicados por Jackson (2010), o genoma de *T. cruzi* contém pelo menos quatro variantes de δ -amastinas e duas β -amastinas por genoma haplóide, sendo que as δ -amastinas foram descritas inicialmente como mais expressas em formas amastigotas. A fim de desvendar as funções das diferentes subfamílias das amastinas são necessárias caracterizações adicionais. Nesse sentido, a determinação da organização genômica, padrão de expressão, localização subcelular em diferentes cepas, e/ou perturbação na expressão destes genes podem fornecer importantes indícios da função dessas proteínas.

Vários recursos para caracterização da função de genes, bastante utilizados em outros organismos, são de aplicação restrita em *T. cruzi* devido às particularidades na biologia molecular deste organismo.

Uma forma bastante usada para estudos de caracterização funcional é a deleção ou silenciamento do gene. Porém, a deleção por recombinação homóloga é bastante complicada para genes multicópias, além de limitar bastante o estudo de genes essenciais. O silenciamento gênico por RNAi em *T. cruzi* não é possível já que alguns componentes da maquinaria enzimática desta via estão ausentes neste organismo. Uma alternativa para deleção seria o uso do sistema CRE/loxP.

A recombinase CRE é capaz de deletar um trecho de DNA flanqueado por loxP. Porém, quando expressa em sua forma nativa, ela pode ser tóxica para o organismo, como descrito em *T. brucei*. Uma alternativa para evitar este problema foi a separação desta proteína em duas partes que se unem na presença de rapamicina tornando-a ativa. Este sistema, chamado DiCre, foi usado com sucesso em leveduras, células murinas e em *Toxoplasma gondii*. A necessidade de adição do ligante para ocorrer a recombinação possibilita a deleção de genes essenciais, uma vez que seria possível induzir e acompanhar a deleção do gene e o fenótipo apresentado pelo organismo. Este sistema viabiliza também o estudo de genes multicópias, inserindo sítios loxP nas pontas do conjunto de genes que se deseja estudar, como é o caso das δ -amastinas. Outra possibilidade seria a retirada de marcas de seleção, que ampliaria o número de modificações possíveis em uma mesma população, já que o número de genes de resistência disponíveis para *T. cruzi* é muito reduzido. Sendo assim, a otimização deste sistema em tripanossomatídeos permitirá grandes avanços nos estudos de genômica funcional nestes protozoários, isto devido a ampla gama de possibilidades de utilização como ferramenta de manipulação da expressão de genes.

Desta forma, propomos estudar a função de subfamílias de amastina pela caracterização de seus padrões de expressão e obtenção de transfectantes. Paralelamente, propomos o desenvolvimento do sistema DiCre em *T. cruzi*,

4 OBJETIVOS

Caracterizar a função de amastinas e desenvolver o sistema de recombinação mediado por recombinase CRE em *Trypanosoma cruzi*.

4.1 OBJETIVOS ESPECÍFICOS:

- Revisar a anotação de amastinas no genoma e reclassificar os genes de amastinas encontrados em *T. cruzi*;
- Superexpressar amastinas:
 - Expressar as duas β -amastinas e δ -Ama40 em fusão com GFP em clones ou cepas diferentes do parasito;
 - Determinar a localização subcelular das amastinas fusionadas a GFP das diferentes subfamílias;
 - Avaliar o fenótipo do parasito expressando as proteínas citadas acima.
- Deletar amastinas
 - Nocautear por recombinação homóloga o gene de δ -Ama40 e as β -amastinas presentes nos cromossomos 26 e 32;
 - Verificar as alterações fenotípicas causadas pelas deleções citadas acima;
- Desenvolver e testar o funcionamento do sistema de recombinação DiCre/*loxP* em *T. cruzi*;

5 MATERIAIS E MÉTODOS

5.1 ANÁLISE DE SEQUÊNCIAS DO GENOMA

As sequências de amastinas foram obtidas a partir dos bancos de dados do genoma de *T. cruzi*, das cepas/ clones CL Brener, Esmeraldo e Sylvio X-10 (TRITRYP DB, 2014). As sequências listadas na

figura 9 foram nomeadas de acordo com a anotação do genoma de CL Brener ou o *contig* ou *scaffold* ID para Sylvio X-10. Todas as sequências codificantes foram traduzidas e alinhadas usando o programa ClustalW (LARKIN *et al.*, 2007). Sequências de aminoácidos provenientes de CL Brener, Esmeraldo, Sylvio X-10, e *Angomonas* sp. (ATCC 30255) foram submetidos a construção da árvore de máxima verossimilhança usando a versão 4.4 do programa SeaView (GOUY *et al.*, 2010) e a árvore filogenética foi construída usando uma α -amastina de *Angomonas* sp como raiz. O programa Weblogo 3.2 foi utilizado para exibir os níveis de conservação de sequências ao longo da proteína (CROOKS *et al.*, 2004). As sequências de aminoácidos de uma amastina de cada subfamília foram utilizadas para prever os domínios transmembrana, utilizando o software Sosui (HIROKAWA *et al.*, 1998).

5.2 CULTIVO

5.2.1 Bactérias

Foram utilizadas para este trabalho bactérias da linhagem DH10B/[Sm^R; F'*proAB*⁺ *lacZ*_M15] (SAMBROOK *et al.*, 1989). Estas foram mantidas em glicerol 50% a -20° C e cultivadas em diferentes meios dependendo do propósito:

- Luria Bertani líquido ou LB (SAMBROOK *et al.*, 1989) (5 g/L de extrato de levedura; 10 g/L de NaCl; 10 g/L de triptona).

- Luria Bertani ágar ou LA (meio LB contendo 15 g/L de ágar).
- SOB (GIBCO-INVITROGEN) (5 g/L de extrato de levedura; 0,6 g/L de NaCl; 0,18 g/L de KCl; 20 g/L de triptona).
- SOC (meio SOB contendo 3,6 g/L de glucose; 0,94 g/L de $MgCl_2$; 1,2 g/L de $MgSO_4 \cdot 7H_2O$).

5.2.2 Parasitos

Foram utilizados para este projeto parasitos da espécie *Trypanosoma cruzi* das cepas e clones CL Brener, Colombiana, clone da cepa Dm28, G, Sylvio X-10 e Y. Estes foram mantidos em meio LIT completo (*Liver Infusion tryptone*), que contém infuso de fígado 1%; NaCl 68,4 mM; KCl 5,4 mM; glucose 5,5 mM; bacto-triptose 0,5%; Na_2HPO_4 56,4 mM; hemina 0,002% suplementado com 10% de soro fetal bovino (Cultilab) e 1 % de solução contendo Penicilina 10000 U e estreptomicina 10000 µg (PenStrep - Sigma). Em algumas seleções de populações resistentes, foi utilizado o meio M199 (Cultilab), suplementado com 10% de LIT completo.

5.3 VETORES

5.3.1 Construídos anteriormente

Os vetores pROCK FEKO_Pur Neo e pODiCre Higo foram construídos previamente em nosso laboratório, com objetivo de testar o funcionamento do sistema DiCre (JULLIEN *et al.*, 2003; 2007) em *Trypanosoma cruzi*.

Para obter pROCK FEKO_Pur Neo, a construção contendo o gene que codifica a resistência a puromicina (PAC) e o gene suicida timidina quinase do vírus da Herpes (HSV-TK) foi retirada do plasmídeo pyrFEKO_Pur, gentilmente cedido por George Cross, da The Rockefeller University, Estados Unidos, e ligada ao vetor pROCKET L19 Neo, como esquematizado na figura 7.

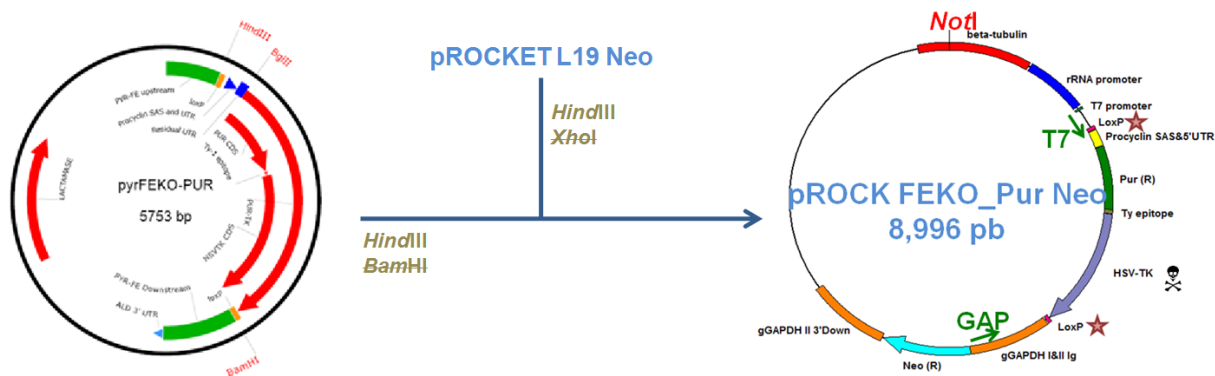


FIGURA 7 – ESQUEMA DE CONSTRUÇÃO DE pROCK FEKO_Pur Neo

FONTE: Adaptado de site do George cross SCAHILL *et al.*, 2008; e o Autor (2014)

NOTA: A figura mostra pROCK FEKO_Pur Neo, um plasmídio contendo PAC (Puromycin N-acetyltransferase) fusionada com HSV-TK (Timidina quinase de Herpes Simplex virus – gene suicida “☠”) flanqueado por sítios *loxP* (★). Toda a construção contendo PAC, HSV-TK e sítios *loxP* são derivados de pyrFEKO-Pur (Scahill et al 2008). pROCK FEKO_Pur Neo, contém sequências regulatórias de genes de *T. cruzi* e uma região de aproximadamente 900 pares de base de um fragmento de β -tubulina após a linearização do plasmídio com *NotI*.

O plasmídio pODiCre Higo foi construído para expressão da enzima CRE recombinase dimerizável em *T. cruzi*. O esquema de construção do vetor está apresentado na figura 8. A construção Cre59, fusionada com o domínio FRB foi retirada do plasmídio pcDNA Cre59 e inserida no pTREX, a jusante da região intergênica de gGAPDH, gerando pTREX Cre59. O trecho contendo Cre59, domínio FRB e gGAPDH IR, foi inserido no plasmídio pcDNA Cre60. Deste, foi retirada toda a construção contendo Cre59, domínio FRB, gGAPDH IR, Cre60 e domínio FKBP12 e inserida em pROCK TK Higo. Os plasmídios pcDNA Cre59 e pcDNA Cre60, foram gentilmente cedidos por Jean-Paul Herman, da Université de la Méditerranée, França.

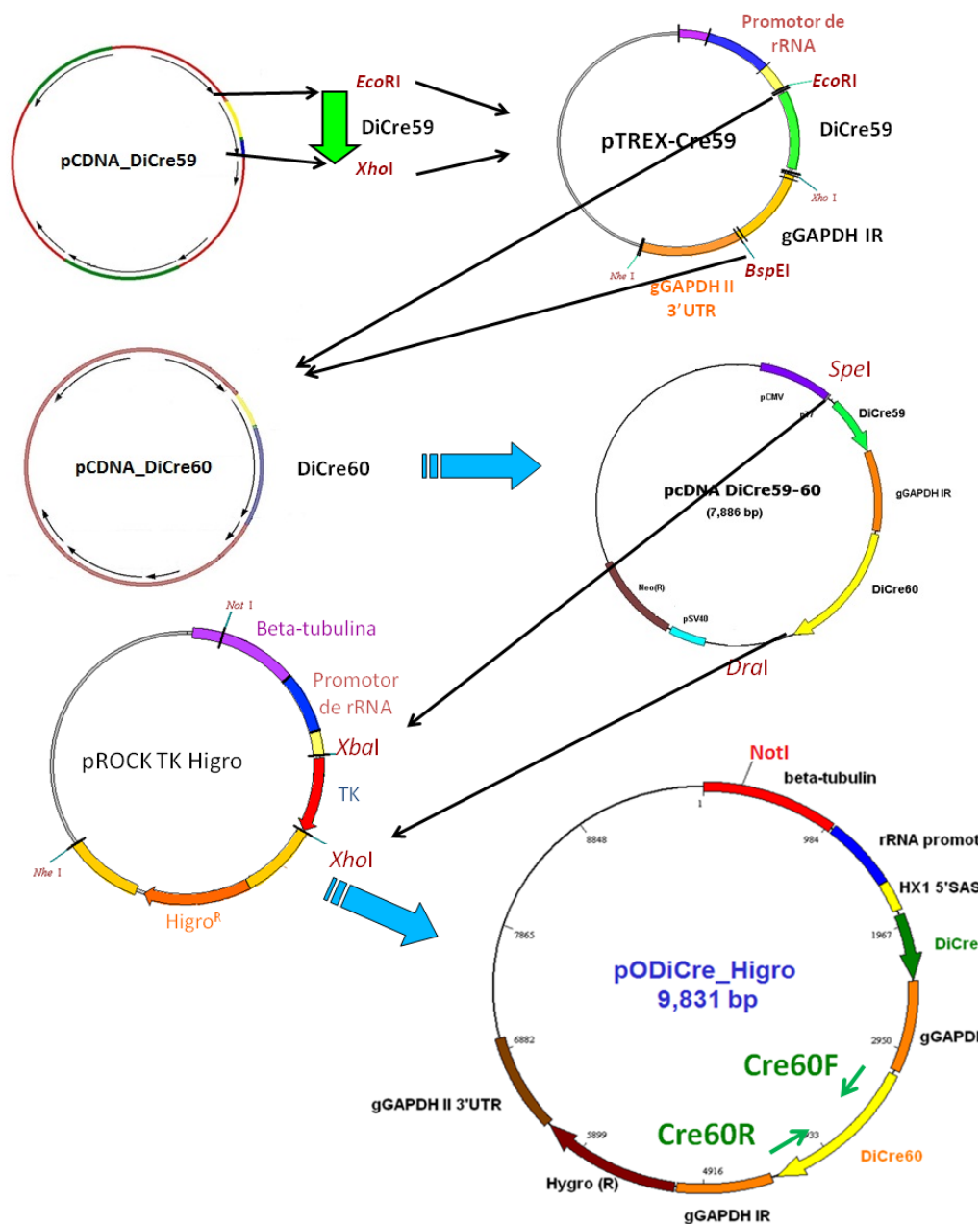


FIGURA 8 – ESQUEMA DE CONSTRUÇÃO DO PLASMÍDIO pODiCre

FONTE: O autor (2014)

NOTA: Para obtenção do vetor pODiCre_Higro a região correspondente a DiCre59 foi retirada do plasmídeo pDNA_DiCre59 (JULLIEN *et al.*, 2003) e inserida no pTREX. Deste plasmídeo, chamado pTREX-Cre59, foi retirada a DiCre59 e a região intergênica de gGAPDH e inserida em pCDNA_DiCre60 (JULLIEN *et al.*, 2003), formando pcDNA DiCre59-60. Deste, foi retirada DiCre59, gGAPDH IR e DiCre60. Toda esta construção foi inserida em pROCK TK Higro, no lugar de TK, gerando, por fim, o pODiCre_Higro

5.3.2 Outros vetores

Os plasmídios que foram utilizados neste trabalho estão descritos na tabela 1.

TABELA 1: PLASMÍDIOS

Plasmídio	Características	Resistência	Referência
pTZ57R/T	Vetor de clonagem TA, LacZ	Ampicilina	FERMENTAS
UP26-DN26	Vetor para deleção por recombinação homóloga do gene da δ -amastina do cromossomo 26 de <i>Trypanosoma cruzi</i> . As regiões UP26 e DN26 foram amplificadas do genoma de <i>T. cruzi</i> das cepas CL Brener e Dm28.	Ampicilina, Puromicina/ Neomicina/ Higromicina	Este trabalho
UP32-DN32	Vetor para deleção por recombinação homóloga do gene da β -amastina do cromossomo 32 de <i>Trypanosoma cruzi</i> . As regiões UP26 e DN26 foram amplificadas do genoma de <i>T. cruzi</i> das cepas CL Brener e Dm28.	Puromicina/ Neomicina/ Higromicina	Este trabalho
pGEM –T Easy	Promotores T7 e SP6 de RNA polimerase em lados opostos; vetor linearizado com tiaminas livres para aumentar a capacidade de clonagem de produtos de PCR	Ampicilina	Promega
pTREX GFP	Vetor de integração ao genoma e expressão da proteína fluorescente GFP. A região codificadora da δ -Ama40 foi amplificada do genoma de <i>T. cruzi</i> das cepas CL Brener.	Ampicilina e Neomicina	GUEVARA <i>et al.</i> , 2005
pTREX Ama40 GFP	Vetor de integração ao genoma e expressão da δ -Ama40 em fusão com GFP. A região codificadora da δ -Ama40 foi amplificada do genoma de <i>T. cruzi</i> das cepas CL Brener.	Ampicilina e Neomicina	Este trabalho
pTREX Ama390 GFP	Vetor de integração ao genoma e expressão da β -amastina 390 em fusão com GFP. A região codificadora da β -amastina 390 foi amplificada do genoma de <i>T. cruzi</i> das cepas CL Brener.	Ampicilina e Neomicina	Este trabalho

pTREX Ama394 GFP	Vetor de integração ao genoma e expressão da β -amastina 394 em fusão com GFP. A região codificadora da β -amastina 394 foi amplificada do genoma de <i>T. cruzi</i> das cepas CL Brener.	Ampicilina e Neomicina	Este trabalho
pTREX 100 Cre Hygro	Vetor de integração ao genoma e expressão da Cre íntegra em fusão com GFP.	Ampicilina e Higromicina	Este trabalho

5.4 INICIADORES

Os iniciadores que foram utilizados neste trabalho estão listados na tabela 2.

TABELA 2: INICIADORES

Nome do iniciador	Sequência (5'→3')
NeoFor	CGACCCTGCAGCCAATATGGGATCG
NeoRev	TCAGAAGAACTCGTCAAGAAGGCG
Cre60For	TCCGGATGGCCCCTAAGAAGAAGAG
Cre60Rev	TCCGGATCTCAGTCCCCATCCTCGA
HygroEcoRI_F	AAAAGAATTTCATGAAAAAGCCTGAACTCAC
HygroEcoRI_R	TTTTGAATTCTATTCTTTGCCCTCGGAC
40Xba_For	TTGTTCTAGAGTAGGAAGCAATG
40Bam_Rev	CGCTGGATCCGAACCACGTGCA
390AvrII_For	CCTAGGAGGATGTCTGAAGAAGAAG
390BglI_Rev	AGATCTCGAGCACAATGAGGCCCA
394Xba_For	TCTAGATGGGCTTCGAAACGCTTGC
394Bam_Rev	GGATCCCCAGTGCCAGCAAGAAGACTG
26UPForHinBa	TATAAGCTTGGATCCACCGCGCCTTCGTTGCTG
26UPRevBgl	ATCCATACGACACAGATCTTCGCCACAC
26DNForXho	TTTCTCGAGGTTTAGCGGTTTTGTGTG
26DNRevXbaI	AACTCTAGACCTCTCCTTCTCAACCTCAG
32UPForHind	TTGAAGCTTTTGCATACTAGGTGTGCGCCTGC
32UPRevBgl	AGTTCCGCAAGATCTCCCTTTGTACACG
32DNForXhoI	CAGCTCGAGTTGTTCATCCGGAACGTGC
32DNRevBaXb	AGCTCTAGAGGATCCAGAAGACCACAGAACTGC
32UPcon	ACGTGTTTCGGCGTCATCGA

5.5 AMPLIFICAÇÃO DO DNA POR PCR

A técnica da reação em cadeia da polimerização (*Polymerase Chain Reaction* - PCR) permite obter um fragmento específico de DNA por meio de sua duplicação em modo exponencial, a partir de uma pequena quantidade de DNA molde. As reações de amplificação foram feitas em termociclador 2720 Thermal Cycler (Applied Biosystems®) e os sistemas de reação continham enzima *Taq* DNA polimerase 1-5 unidades por reação de 20-100 µL, tampão de PCR FERMENTAS® 10X *Taq* Buffer + KCl [100 mM Tris-HCl (pH8,8); KCl 500mM; Nonidet P40 0,8% (v/v)] ou 10X *Taq* Buffer + [(NH₄)₂SO₄ (750 mM Tris-HCl (pH8,8); 200 mM (NH₄)₂SO₄; 0,1% (v/v) Tween 20)], dNTP 0,2 mmol/L, MgCl₂ 1,5 mmol/L e 5-10 pmol de oligonucleotídeos iniciadores (AZEVEDO *et al.*, 2003-adaptado). Os parâmetros de ciclos de amplificação variaram conforme os oligonucleotídeos utilizados.

Para amplificação de cassetes para nocaute foram feitas 10 reações de 100 µL nas mesmas condições descritas acima. Após verificar a amplificação por eletroforese em gel de agarose, a reação foi tratada com *DpnI* a 37° C overnight, para clivagem do DNA plasmidial. Em seguida o produto de amplificação foi precipitado com 2,5 volumes de etanol 96% e 300 mM de Acetato de sódio a -20° C por 1 hora, centrifugado por 20 min a 25000 xg, lavado com etanol 70% e adicionado de água estéril. A concentração de DNA foi medida em absorbância a 260 nm e foi utilizado de 7 a 10 µg para transfecção.

5.6 RESTRIÇÃO ENZIMÁTICA DE DNA

Para obtenção de DNA para ligação, uma unidade de enzimas de restrição (FERMENTAS®) foram utilizadas para digestão de 0,5 – 1 µg de DNA em um volume final de 50 µL, durante no mínimo 3 horas, na presença do tampão recomendado e temperatura adequada.

Para digestão de DNA genômico para *Southern blot*, foi usado 1 a 5 unidades de enzimas de restrição (FERMENTAS[®]) para digerir 7 µg de DNA, em volume final de 30 µL, durante a noite, na presença do tampão recomendado e em temperatura adequada.

Após a clivagem o DNA foi precipitado com isopropanol 0,6 volume, lavado com etanol 70% e adicionado de água ultrapura.

5.7 ELETROFORESE DE DNA

Após a extração de DNA, digestão, ou reação de PCR, a amostra foi submetida a eletroforese para avaliação do resultado. A eletroforese de DNA foi feita em gel de agarose horizontal como descrito por SAMBROOK e colaboradores (1989). O tampão utilizado foi TBE 1X (Tris-HCl 8,9 mM; ácido bórico 8,9 mM e EDTA 0,2 mM; pH 8,0). As amostras foram diluídas em tampão F.E.F Suds (azul de bromofenol 0,025% (p/v); xilenocianol 0,025% (p/v); SDS 0,1%(p/v); Ficoll[®] 20% em tampão T₁₀E₁). Após o tempo de corrida, que pode variar dependendo da amperagem utilizada, o DNA foi corado com solução de brometo de etídeo (0,05%) por 15 min e visualizado sob luz ultravioleta (302 nm) em transiluminador UV. As imagens foram obtidas utilizando o EC3[™] Imagem System - UVP BioImaging Systems.

5.8 PURIFICAÇÃO DE FRAGMENTOS

Uma corrida nas mesmas condições descritas no item 5.7 foi realizada em maior volume para obtenção do fragmento desejado para purificação. A banda que corresponde ao perfil esperado foi cortada do gel e purificada por kit *Illustra™ GFX™ PCR DNA and Gel Band Purification Kit* (GE Healthcare). A eficiência de purificação foi analisada por eletroforese.

5.9 REAÇÕES DE LIGAÇÃO

As reações de ligação foram feitas com fragmentos purificados ou vetor comercial. Os diferentes insertos e vetores foram ligados numa razão molar de 1:5. Nas reações de ligação foram adicionados tampão de ligação e 0,5 U de T4 DNA ligase. O sistema foi incubado a 16° C durante a noite.

5.10 PREPARO DE CÉLULAS TERMOCOMPETENTES E TRANSFORMAÇÃO BACTERIANA POR CHOQUE TÉRMICO

Inicialmente foi feito um pré-inóculo com *Escherichia coli* DH10B, linhagem resistente a tetraciclina, com 4 mL de LB suplementado com 100 µg/mL de tetraciclina, sendo incubado durante a noite a 37 °C sob agitação. Em seguida um inóculo foi preparado com 100 mL de LB suplementado com 100 µg/mL de tetraciclina sendo incubado a 37 °C sob agitação até atingir DO₆₀₀ 0,5.

A cultura foi transferida para tubos plásticos e estes foram mantidos no gelo por 20 min, com posterior centrifugação a 2000 xg por 8 min a 4 °C. O sobrenadante foi descartado e o sedimento misturado a 30 mL de uma solução de MgCl₂ 100 mM. A suspensão foi centrifugada nas mesmas condições anteriores. O sobrenadante foi novamente descartado e o sedimento misturado a 30 mL de CaCl₂ 100 mM. Após incubação em banho de gelo por 30 min, o mesmo foi centrifugado a 2000 xg por 8 min a 4°C, o sobrenadante descartado e sedimento ressuspenso em 2 mL de uma solução de CaCl₂ 85 mM em Glicerol 15% e aliqotado 200 µL em tubos de 1,5 mL a -80°C ou transformadas em seguida.

Alíquotas de bactérias *E. coli* termocompetentes foram misturadas com o produto de ligação ou plasmídeo, mantidas em banho de gelo por 30 min e transformadas por choque térmico a 42 °C por 90 segundos, seguidas de banho de gelo por 3 min. A este material foram adicionados 500 µL de meio LB ou SOC sendo incubado por 1 hora a 37 °C. Em seguida, 200 µL da cultura transformada foram

plaqueados em meio LA suplementado com 100 µg /mL de ampicilina. No caso do pGEM-T Easy, o meio foi suplementado ainda com 0,5 µM de IPTG e 40 µg/mL de X-gal e as placas incubadas por 16 horas a 37 °C. Os clones foram selecionados por PCR de colônia, utilizando os iniciadores do inserto ou do vetor e uma colônia em cada reação, ou por perfil de restrição. O DNA foi posteriormente extraído e a presença dos insertos confirmada adicionalmente por restrição.

5.11 OBTENÇÃO DE DNA PLASMIDIANO

A purificação de plasmídios a partir de *E. coli* em pequena escala foi baseada no método da lise alcalina segundo SAMBROOK *et al.*, (1989) com algumas modificações.

Bactérias *E. coli* DH10B, contendo o plasmídio de interesse foram cultivadas em 4 mL de meio LB adicionado de antibióticos seletivos. Esta cultura foi transferida para tubos plásticos de 2 mL, as células foram coletadas por centrifugação (24000 xg, 2 min) e adicionadas de 250 µL de solução GET (Tris-HCl 25 mmol/L pH 8,0; glucose 50 mmol/L; EDTA 10 mmol/L) contendo 10 µg/mL de RNase. Em seguida as células foram lisadas com adição de 250 µL de uma solução de lise (NaOH 0,2 mol/L e SDS 1% (m/v)). Em seguida, as proteínas, o DNA cromossômico, os restos celulares e o SDS foram precipitados com adição de 200 µL de acetato de potássio (3 mol/L, pH 4,8). Foram adicionados mais 100 µL de clorofórmio:álcool isoamílico (24:1) e após 5 min as amostras foram centrifugadas (24000 xg, 5 min) e o sobrenadante coletado. O DNA plasmidiano foi precipitado com isopropanol 0,6 volume, lavado com 500 µL de etanol 70%, seco e reidratado em 30 µL de água ultrapura. O DNA purificado foi analisado por eletroforese em gel de agarose 1%.

Para obtenção de plasmídio em larga escala, foi realizada uma MIDI PREP com o *Kit QIAGEN Plasmid Plus Midi*.

5.12 EXTRAÇÃO DE DNA GENÔMICO

Para extração de DNA genômico, a cultura de *T. cruzi* em fase exponencial foi centrifugada a 2500 xg por 10 min a temperatura ambiente. Logo após, o sobrenadante foi descartado e as células foram lavadas com PBSG (NaH₂PO₄ 75 mM, NaCl 60 mM e Glucose 10%). Após nova centrifugação, o sedimento foi homogeneizado em tampão de lise/digestão (NaCl 100 mM, Tris-Hcl 10 mM pH 8,0; EDTA 25 mM pH 8,0; SDS 0,5%, proteinase K 100 µg/mL; RNase 10 µg/mL) a uma concentração de 5x10⁸ parasitos por mL e foi incubado a 37 °C durante a noite. Em seguida o DNA foi extraído com 1 volume de fenol: clorofórmio: álcool isoamílico (25:24:1). A fase aquosa foi transferida para outro tubo, e o DNA foi novamente extraído, agora com clorofórmio: álcool isoamílico (24:1) a fim de obter uma amostra mais pura. A fase aquosa foi novamente transferida para outro tubo, e o DNA foi precipitado com 0,6 volumes de isopropanol, lavado com etanol 70%, seco e reidratado em água ultrapura.

5.13 TRANSFEÇÃO DE PARASITOS

Epimastigotas na fase exponencial de crescimento foram lavados em tampão de eletroporação (KCl 120 mM, CaCl₂ 0,15 mM, K₂HPO₄ 10 mM, Hepes 25 mM, EDTA 2 mM pH 8,0 e MgCl₂ 5mM) a 4 °C. Em seguida os parasitos foram misturados com o mesmo tampão para obter concentração final de 1x10⁸ células/mL. Alíquotas de 0,4 mL (4 x 10⁷ parasitos) desta suspensão de células foram misturados a 50 µg de cada plasmídeo previamente linearizado por digestão, em um volume máximo de 50 µL em cubetas de eletroporação 0,2 cm (Bio-Rad). Os parasitos foram eletroporados utilizando o aparelho Gene Pulser Xcell™ Electroporation System (Bio-Rad). Foi realizada a eletroporação utilizando 0,45 KV, capacitância de 500 µF e 800 ohms, tendo sido aplicados 2 pulsos, com intervalos de 30 segundos entre eles.

Após a eletroporação, os parasitos foram transferidos para garrafas de cultura estéreis contendo 5 mL de meio LIT suplementado com 10% SFB (soro fetal bovino) e incubadas a 25-27 °C. No dia seguinte à eletroporação foram adicionados 200 a 500 µg/mL de G418 (Sigma) ou de Higromicina (Sigma) à cultura. Para acompanhar a seleção dos parasitos resistentes, uma amostra passou pelos mesmos procedimentos que as populações transfectadas com a diferença de não conter DNA plasmidiano. Esta população recebeu as mesmas doses de antibióticos e foi diluída nas mesmas proporções que as transfectadas. Considerou-se que estavam selecionados os parasitos resistentes quando nesta população não foram encontrados indivíduos viáveis.

5.14 CLONAGEM DOS PARASITOS TRANSFECTADOS

Para a obtenção de clones dos parasitos geneticamente modificados, culturas em fase exponencial de crescimento foram submetidas a um passo de diluição limitante até a concentração final de 0,5 parasitos/mL em meio LIT suplementado com 10% de soro fetal bovino e os antibióticos de seleção. Em seguida, a cultura foi distribuída em placa de 24 poços, a 1 mL por poço. As placas foram incubadas a 25-27 °C em câmara úmida até que foi observado crescimento em alguns poços da placa. Os parasitos foram transferidos para garrafas de cultura contendo meio LIT suplementado com 10% de soro fetal bovino e os antibióticos de seleção.

5.15 CURVA DE RESISTÊNCIA A PUROMICINA

A curva de resistência a puromicina foi desenvolvida com a população resistente a G418 transfectada com pROCK FEKO_Pur Neo para verificar se o gene de resistência a puromicina estava sendo expresso em *T. cruzi*.

Para isto, a cultura em fase logarítmica foi diluída para 1×10^6 parasitos/mL na ausência ou presença de 2 µg/mL de puromicina, seu crescimento foi acompanhado

diariamente através de contagem em câmara de Neubauer. Como controle de sensibilidade ao antibiótico, o mesmo procedimento foi realizado com a cepa selvagem. As culturas foram diluídas ao atingir a densidade de 1×10^7 parasitos /mL.

5.16 TESTE DE RECOMBINAÇÃO POR INDUÇÃO COM RAPAMICINA

Após a obtenção de clones de *T. cruzi* contendo pROCK FEKO_Pur Neo e pODiCre Higro foi realizado o teste de recombinação por DiCre induzida por rapamicina. Para isto, os clones duplo resistentes mantidos em fase exponencial foram diluídos a 5×10^6 parasitos/ mL e a eles foi adicionada rapamicina em concentrações variando entre 50 nM e 1 μ M e foram mantidos a 25-27 °C por diferentes tempos de indução. Em seguida parte da cultura foi utilizada para extração de DNA genômico e o restante foi lavado e mantido em cultura. Para tempos de indução curtos, a cultura foi lavada e mantida por alguns dias a 25-27 °C para obtenção de número suficiente de células para extração de DNA genômico. Foram utilizados como controle, nos tempos de indução e concentração de rapamicina mais altos, os clones contendo apenas um dos plasmídios (pROCK FEKO_Pur Neo ou pODiCre Higro), além da cepa CL Brener selvagem. As amostras de DNA genômico obtidas dos testes foram digeridas para ensaio de RFLP (item 5.22).

Em um segundo ensaio, a fim de obter populações clonais para o perfil de recombinação, a população duplo-resistente contendo pROCK FEKO_Pur Neo e pODiCre Higro para indução por 48 horas a 1 mM de rapamicina e em seguida foi realizada clonagem por diluição seriada em placas de 24 poços, sendo que um poço foi usado para controle de crescimento da cultura, contendo 100 parasitos /mL. Os demais poços continham 1 mL de uma suspensão diluída a 0,5 parasitos /mL. Foram feitas quatro placas para a população duplo-resistente induzida, sendo duas sem o indutor e duas com 1 mM de rapamicina. A população não induzida foi utilizada como controle, e desta, foram feitas duas placas para clonagem. Os demais controles foram as populações contendo somente pROCK FEKO_Pur Neo ou

somente pODiCre Higro e o parasito selvagem do clone CL Brener. Os clones foram expandidos para extração de DNA genômico, que foi utilizado para ensaio de RFLP (item 5.22).

5.17 OBTENÇÃO DE EXTRATO PROTÉICO TOTAL DE EPIMASTIGOTAS DE *T. CRUZI*

Cerca de 8×10^7 parasitos em fase exponencial de crescimento foram coletados do meio LIT por centrifugação a 2500 xg durante 10 min e lavados com tampão PBS gelado. Estes parasitos coletados foram lisados em 200 µL de tampão de amostra SDS-PAGE 1x (Tris-HCl 50mM, pH 6,8, Glicerol 10%, SDS 2%, DTT 5mM, azul de bromofenol 0,1%). As amostras foram mantidas a -20° C até a sua utilização.

5.18 OBTENÇÃO DA FRAÇÃO ENRIQUECIDA DE MEMBRANA DE EPIMASTIGOTAS DE *T. CRUZI*

A fim de se obter fração enriquecida de membrana de epimastigotas de *T. cruzi*, foi realizado o fracionamento celular. Aproximadamente 1×10^9 parasitos/mL foram lavados uma vez em PBS 1x e foram misturados em 3 ml de tampão de extração de fração de membrana (Hepes 20 mM; KCl 10 mM; $MgCl_2$ 1,5 mM; sacarose 250 mM; DTT 1 mM; PMSF 0,1 mM). Após a adição do tampão o homogeneizado foi armazenado a - 20° C por 48 horas. Em seguida os parasitos foram lisados com cinco ciclos de congelamento no nitrogênio líquido e descongelamento em água a 37° C. O lisado foi centrifugado a 2500 xg por 10 min e o sobrenadante (proteína total - T) foi submetido à ultracentrifugação a 105.000 xg, 4° C por 1 hora. O sobrenadante desta centrifugação contendo a fração de proteínas solúveis (denominada fração S) foi coletado e armazenado a - 80° C. O precipitado contendo a fração enriquecida com proteínas de membranas (denominada fração I)

foi homogeneizado no mesmo tampão de extração de proteína de membrana e armazenado nas mesmas condições do sobrenadante até a sua utilização (BERGER; KRETZLER, 2002). As concentrações proteicas das amostras foram determinadas segundo o método quantitativo de Bradford (1976) através da absorbância obtida no espectrofotômetro a 600 nm utilizando concentrações conhecidas de BSA como padrão.

5.19 SDS-PAGE DO EXTRATO PROTÉICO TOTAL E DAS FRAÇÕES ENRIQUECIDAS

Cerca de 20 µL do extrato proteico total ($0,8 \times 10^7$ parasitos/mL) foram submetidos à separação eletroforética em gel SDS-PAGE 12%. A corrida foi realizada em tampão de eletroforese para SDS-PAGE (Tris 25 mM, glicina pH 8,3; 0,25 M, SDS 0,1%) a 100V por 2 horas.

Para as amostras do fracionamento celular de proteínas totais, frações enriquecidas citoplasmáticas e de membrana, foram utilizados 1 µg de proteínas para os géis de coloração com prata e 20 µg para análises por *western blot*. Cada amostra foi preparada com o tampão SDS-PAGE 1x.

5.19.1 Coloração com Azul de Coomassie

Os géis foram corados pela imersão em uma solução de azul de coomassie 0,1% em metanol: ácido acético: água (5:2:2) por 3 h, seguida de descoramento em solução de metanol 30% e ácido acético 10%.

5.19.2 Coloração com Prata

Depois da separação eletroforética, os géis foram lavados por 5 min com água destilada, e fixados em metanol 50% por 20 min. Foi feita outra lavagem de 10

min e sensibilização com glutaraldeído 10% por 20 min. Em seguida foram realizadas quatro lavagens de 10 min com água destilada para a coloração com uma solução de nitrato de prata (NaOH 0,2%; NH₄OH 0,5 M, AgNO₃ 60 mM) por 20 min. Posteriormente foi realizada duas lavagens de 5 min com água destilada e revelação com solução reveladora (ácido acético 4 mM; formaldeído 0,37%) até o aparecimento das bandas. Esta reação foi interrompida utilizando a solução de parada (ácido acético 1%) por 30 min. O gel foi armazenado em uma solução de armazenamento (etanol 25%; glicerol 1,5%).

5.20 WESTERN BLOT DE PROTEÍNAS DE EPIMASTIGOTAS DE *T. CRUZI*

Nos ensaios de *western blot*, 20 µg das proteínas do fracionamento celular ou 20 µL do extrato protéico total separados por eletroforese, foram transferidas para a membrana de nitrocelulose Hybond-C (Amersham) a 4° C, 100 V, durante 1 hora e 20 min em tampão de transferência (Tris 25 mM, glicina pH 8,3; 0,2 5M, SDS 0,1%, Metanol 20%). Em seguida, a membrana foi bloqueada em solução PBS/T (PBS 1x com 0,05% de tween 20) com 5% de leite em pó desnatado a 4° C durante a noite. Após o bloqueio, a membrana foi lavada 3x por 5 min a 4° C e foi incubada com o anticorpo primário. Em seguida foi realizada 3 lavagens de 5 min com PBS/T e a membrana foi incubada com anticorpo secundário conjugado a peroxidase. Posteriormente foram feitas 3 lavagens de 5 min com PBS/T e a membrana foi revelada em filme GE por quimioluminescência usando *Kit ECL + PLUS* (GE heathcare), seguindo instruções do fabricante.

5.21 ELETROFORESE DE DNA EM CAMPO PULSADO (PFGE)

A cultura de *T. cruzi* em fase exponencial foi centrifugada a 2500 xg por 10 min a temperatura ambiente. Logo após, o sobrenadante foi descartado e as células foram lavadas com PBSG (NaH₂PO₄ 75 mM, NaCl 60 mM e Glucose 10%). Após

nova centrifugação, o sedimento foi misturado cuidadosamente a PBSG em a uma concentração de 10^7 parasitos por 50 μL . Igual volume de agarose de baixo ponto de fusão (LMP - *Low Melting Point*) 2%, à temperatura de aproximadamente 40 °C, foi adicionado à suspensão de parasitos. Esta mistura foi, então, homogeneizada e distribuída em moldes com volume de 100 μL com concentração final de parasitos de 1×10^5 / μL .

Após solidificação, cerca de 10 min a 4 °C, os blocos foram mergulhados em solução ESP (EDTA 0,5 M, Lauroyl sarcosine 1% e Pronase 0,5 mg/mL), e mantidos por 48 horas a 37 °C. Após lavagem em EDTA 0,5 M pH 8,0, os blocos foram armazenados a 4 °C nesta solução até o momento da eletroforese (ENGMAN *et al.*, 1987).

Os blocos contendo cromossomos dos parasitos foram submetidos à PFGE ao lado de um padrão comercial de peso molecular (cromossomos de *Hansenula wingei* ou *Saccharomyces cerevisiae*, Bio Rad). As bandas cromossômicas dos parasitos foram separadas através de eletroforese de campo pulsado (PFGE) conforme descrito por Cano *et al.* (1995), com algumas modificações.

Os géis de PFGE foram preparados com agarose 0,8% em TBE 0,5X (25 mM Tris-borato/0,5 mM EDTA, pH 8,0). As corridas eletroforéticas foram realizadas por 135 horas com voltagem constante (80 V) a 13 °C em aparelho *Gene Navigator System* (Pharmacia). Foram aplicados pulsos homogêneos (N/S, L/O) de 90 s por 30 h, 120 s por 30 h, 200 s por 24 h, 350 s por 25 h e 800 s por 26 h, com interpolação. Após a separação cromossômica os géis foram corados com solução de brometo de etídeo (0,5 $\mu\text{g/mL}$) e fotografados. Em seguida, o DNA foi transferido para uma membrana Hybond-N+ (AMERSHAM) pelo protocolo de *Southern blot* (item 5.22) e hibridizado com sonda apropriada marcada radioativamente (item 5.24).

5.22 TESTE DE POLIMORFISMO DO COMPRIMENTO DO FRAGMENTO DE RESTRIÇÃO (RFLP) E SOUTHERN BLOT

Foi extraído DNA genômico de culturas de *T. cruzi* selvagem, transfectados ou testes. Estes DNAs foram digeridos com enzima apropriada durante a noite, em tampão e temperatura adequados. Depois de confirmada a digestão completa por eletroforese em agarose 1%, as amostras foram aplicadas em agarose 0,8% e corridas em TBE 1X a 25V por 13 horas. O gel foi corado em brometo de etídio e fotografado com régua para obtenção das posições das bandas do marcador de peso molecular 1 Kb (Fermentas).

Para tornar possível a transferência do DNA para uma membrana, o gel foi tratado de acordo com o protocolo descrito por SOUTHERN, 1975 com algumas modificações. Primeiramente o DNA foi depurinado com uma solução 125 mM de HCl por 10 min (30 min para géis de PFGE). Em seguida, foi incubado com solução desnaturante (NaOH 0,5 M, NaCl 1,5 M) por 30 min em agitador orbital. Na sequência, foi neutralizado em solução contendo Tris-HCl 0,5 M, pH 7,5 e NaCl 1,5 M por duas vezes de 15 min. Por fim foi incubado por 10 min em SSC 10x. Foi então montado um sistema para transferência do DNA para membrana Hybond-N+ (AMERSHAM) por capilaridade com solução SSC 10X. O tempo de transferência pode variar de 1 a 3 dias. Após a transferência, foi feito o crosslinking do DNA na membrana em luz UV por 3 min. Em seguida a membrana foi mantida a -20 °C até sua hibridização.

5.23 NORTHERN BLOT

Culturas de parasitos transfectados e selvagem, em fase exponencial de crescimento, foram centrifugadas a 2500 xg por 10 min a temperatura ambiente. Logo após, o sobrenadante foi descartado e as células foram lavadas por duas vezes em PBS. O RNA total de aproximadamente 5×10^8 parasitos foi extraído

utilizando o RNeasy Mini Kit (Qiagen) seguindo os procedimentos recomendados pelo fabricante. O método de extração consiste em lise na presença de isotiocianato de guanidina e cromatografia de afinidade, onde RNAs maiores que 200 pb ligam seletivamente a membrana de sílica-gel.

Os géis de agarose para análise das amostras de RNA foram preparados na concentração 1,5 % em tampão MOPS 1X/EDTA (MOPS 0,02 M; Acetato de sódio 5mM; EDTA 1 mM, pH 7,0) e formaldeído 2 %. Foram adicionados 25 µL de tampão de amostra [0,75 mL de Formamida deionizada; 0,15 mL de tampão MOPS 10X; 0,24 mL de Formaldeído; 0,1 mL de Água; 0,1 mL de Glicerol; 0,08 mL de Azul de Bromofenol a 10 % (p/v)] em 10 µg de amostra e incubado a 65 °C por 15 min. Em seguida, as amostras foram aplicadas no gel. Após a corrida, o gel foi corado em brometo de etídeo e analisado em transiluminador de UV. Para a transferência do RNA, o gel foi tratado com solução NaOH 50 mM por 20 min, e em SSC 2X por 40 min. Os RNAs foram transferidos para membrana Hybond-N+ (AMERSHAM) por capilaridade em solução SSC 10X.

5.24 MARCAÇÃO DE SONDAS RADIOATIVAS E HIBRIDIZAÇÃO

Aproximadamente 50 ng de fragmentos de DNA correspondente ao DNA molde para sonda foram radioativamente marcados pelo método "random primer" com 50 µCi de [α -32P]dCTP utilizando 5 µL da solução de hexanucleotídeos randômicos como iniciadores e dNTPs não marcados 1 U enzima *Klenow fragment*. Após 1 hora de marcação a 37° C, as reações foram bloqueadas com 0,05 M de EDTA pH 8,0 e a sonda foi desnaturada com 0,2 M de NaOH e aquecimento a 95° C por 5 min.

Antes da incubação com as sondas, as membranas foram submetidas à pré-hibridização em aproximadamente 30 ml de solução de pré-hibridização (SSC 5X, SDS 1%, Denhardt's 5X e 100 µg/mL de DNA de esperma de salmão desnaturado) em frascos apropriados mantidos em fornos giratórios a 60°C por 1 hora. Após

prévia desnaturação por fervura, as sondas marcadas foram adicionadas a 10 ml de solução de pré-hibridização. As membranas foram expostas a essa solução para hibridização e mantidas a 60°C por aproximadamente 18 horas.

Após hibridização, as membranas foram lavadas uma vez a frio e duas vezes por 45 min em forno giratório com solução de lavagem (SSC 2X e SDS 0,2%) a 60°C. Após a lavagem as membranas foram seladas e expostas em filme de raios X (filme de raios-X, Kodak X-Omat TM) a -70 °C em cassete e reveladas em diferentes tempos de exposição.

5.25 MICROSCOPIA CONFOCAL E IMUNOFLUORESCÊNCIA

Os parasitos expressando amastinas fusionadas com proteínas fluorescentes foram lavados duas vezes com PBS suplementado com 3% de soro fetal bovino e fixados com paraformaldeído 4% em PBS a temperatura ambiente por cerca de 10 min. Posteriormente as amostras foram novamente lavadas duas vezes em PBS suplementado com 3% de SFB e tratadas por 15 min a 37 °C e lavadas mais duas vezes em PBS. Em seguida, o DNA das células foi corado por 10 min a temperatura ambiente com 0,1 µg/mL DAPI (4',6-diamidino-2-phenylindole, Sigma) e montadas sobre lâminas de vidro contendo 4 µL de glicerol 90% em tampão Tris-HCl 50 mM; pH 9,0.

Para as análises de co-localização, parasitos expressando δ -Ama40::GFP foram fixados durante 20 min em paraformaldeído (PFA) a 4% em PBS, à temperatura ambiente. Parasitos aderidos a lamínulas contendo poli-L-lisina (Sigma) foram permeabilizados com 0,1% de Triton X-100 em PBS durante 2 min, bloqueados com 4% de BSA em PBS, durante 1 hora. Em seguida foram incubados com o anticorpo primário [anticorpo policlonal de coelho α -PEPCK (fosfoenolpiruvato carboxiquinase), Curitiba, Brasil, gentilmente cedido pelo Prof. Dr. Stenio Frago, Instituto Carlos Chagas] em solução (4% de BSA em PBS) de bloqueio por 1 hora seguido de incubação com IgG anti-coelho secundário conjugado com Alexa 546.

Amostras também foram coradas com 0,1 µg/mL DAPI (4',6-diamidino-2-phenylindole, Sigma), à temperatura ambiente por 5 min e as lâminas foram montadas para microscopia confocal.

5.26 CURVA DE CRESCIMENTO

Epimastigotas de *T. cruzi* dos clones CL Brener ou Sylvio X-10 expressando ou não amastinas em fusão com GFP foram mantidos em fase logarítmica por pelo menos duas passagens e em seguida inoculados a $0,5 \times 10^6$ parasitos por mL. A partir do dia seguinte, estas culturas foram contadas em câmara de Neubauer a cada 24 horas, por sete dias consecutivos. Nos dias 0 (inóculo), 3, 6, 9 e 12, a fluorescência dos parasitos em cultura foi acompanhada por citometria de fluxo.

5.27 CITOMETRIA DE FLUXO

Epimastigotas de *T. cruzi* expressando proteínas fluorescentes tiveram sua fluorescência monitorada por citometria de fluxo. Aproximadamente 10^6 parasitos crescendo em fase log, ou provenientes da curva de crescimento, foram fixadas durante 20 min em paraformaldeído (PFA) a 4% em PBS, à temperatura ambiente, centrifugados por 5 min a 2500 xg e adicionados de 1 mL de PBS. Estas amostras foram passadas em citômetro de fluxo FACSCalibur BD e avaliadas usando filtro FL1 (530/30 nm), que capta fluorescência verde, emitida em 509 nm. Como controle, a fluorescência dos parasitos selvagens foi medida e considerada como linha base. Para as demais amostras, somente os eventos com fluorescência medida à direita da linha base foram considerados. As análises comparativas de fluorescência foram realizadas pelo software Cyflogic.

6 RESULTADOS E DISCUSSÃO

6.1 PARTE 1: CARACTERIZAÇÃO FUNCIONAL DAS NOVAS SUBFAMÍLIAS DE AMASTINAS

6.1.1 Conservação, estrutura e organização genômica das subfamílias de amastinas de *T. cruzi*

Hoje sabemos que as amastinas, inicialmente descritas como proteínas de superfície mais expressas em amastigota de *T. cruzi*, apresentam uma grande variabilidade, são expressas em outras fases do ciclo biológico em *Leishmania* spp. e estão presentes em outros tripanossomatídeos (JACKSON, 2010; ROCHETTE *et al.*, 2005). Esta família tão diversa, com tal localização, e expressão diferencial, pode apresentar funções importantes para a biologia do parasito, como em sua interação com o hospedeiro ou vetor, na diferenciação, infectividade ou em outros processos. O conhecimento detalhado destas proteínas pode trazer ideias para novos tratamentos ou vacinas visando a diminuição do número de casos ou melhorando as condições de vida dos pacientes já acometidos pela doença.

Com o grande avanço na geração de dados de genoma conquistado na última década, é possível iniciar os estudos desta singular família de proteínas através da análise das sequências de genes que as codificam. Utilizando a classificação de Jackson (2010), porém avaliando somente as amastinas presentes em *T. cruzi*, propomos uma nova divisão das subfamílias de amastinas. A partir da construção de uma árvore filogenética utilizando os 12 genes que estão anotados como amastinas no genoma de CL Brener (sendo seis genes por genoma haplóide), juntamente com os genes anotados no genoma do clone Sylvio X-10 e da cepa Esmeraldo (

figura 9), propõe-se que as amastinas de *T. cruzi* sejam divididas em duas subfamílias. As δ -amastinas, apresentadas em fundo verde, estão subdivididas em

dois grupos, no maior ramo estão as amastinas descritas por Teixeira *et al.*, (1994), cujos genes estão agrupados em conjunto no genoma e intercalados pelo gene de tuzina. No ramo menor das δ -amastinas, encontramos um grupo distinto de amastinas de cópia única no genoma, que apresenta cerca de 50% de similaridade com as demais δ -amastinas, e estão acompanhados por um pseudogene de tuzina. Este subgrupo de δ -amastinas foi denominado δ -Ama40/50.

Uma maior divergência foi encontrada para as β -amastinas, que apresentam de 15 a 35% de similaridade com as δ -amastinas. Os dois ramos destacados em fundo azul e amarelo (figura 9b) agrupam genes de cópia única, que diferem bastante entre si, de 17 a 25% de similaridade. Semelhante ao observado em *Leishmania* sp., estes genes, que foram aqui subdivididos em β 1-amastinas e β 2-amastinas, estão arranjados em sequência (em *tandem*) no genoma de CL Brener e Sylvio X-10.

A fim de evidenciar as diferenças e semelhanças entre estas proteínas, um membro de cada um destes grupos destacados pela árvore filogenética foi escolhido para um alinhamento de aminoácidos mostrado na figura 10a. O alinhamento com todas as sequências citadas na figura 9a foi usado para gerar um weblogo, que mostra a similaridade dos aminoácidos nestas proteínas (figura 10b). Apesar de apresentarem alta variabilidade na sequência de aminoácidos, é mantido o caráter hidrofóbico das regiões transmembrana, que estão sublinhadas em vermelho na figura 10b. Os aminoácidos pertencentes à assinatura das amastinas (C-[IVLYF]-[TS]-[LF]-[WF]-G-X-[KRQ]-X-[DENT]-C), encontrados anteriormente entre as posições 52 e 62 (ROCHETTE *et al.*, 2005), estão cinco aminoácidos deslocados a frente, entre as posições 57 e 67 e encontram-se altamente conservados.

(A)

Identificador de sequência	Cepas de <i>T. cruzi</i>	Localização genômica	Classificação
Tc00.1047053509965.390	CL Brener Esmeraldo-like	cromossomo <i>In silico</i> - TcChr32-S	β1-amastina
Tc00.1047053509051.20	CL Brener Non-Esmeraldo-like	cromossomo <i>In silico</i> - TcChr32-P	β1-amastina
Tc00.1047053509965.394	CL Brener Esmeraldo-like	cromossomo <i>In silico</i> - TcChr32-S	β2-amastina
Tc00.1047053511497.19	CL Brener Non-Esmeraldo-like	cromossomo <i>In silico</i> - TcChr32-P	β2-amastina
Tc00.1047053507739.120	CL Brener Esmeraldo-like	cromossomo <i>In silico</i> - TcChr34-S	δ-amastina
Tc00.1047053506437.10	CL Brener Non-Esmeraldo-like	cromossomo <i>In silico</i> - TcChr34-P	δ-amastina
Tc00.1047053507485.150	CL Brener Esmeraldo-like	cromossomo <i>In silico</i> - TcChr34-S	δ-amastina
Tc00.1047053506437.30	CL Brener Non-Esmeraldo-like	cromossomo <i>In silico</i> - TcChr34-P	δ-amastina
Tc00.1047053507485.130	CL Brener Esmeraldo-like	cromossomo <i>In silico</i> - TcChr34-S	δ-amastina
Tc00.1047053509289.10	CL Brener Non-Esmeraldo-like	cromossomo <i>In silico</i> - TcChr34-P	δ-amastina
Tc00.1047053511071.40	CL Brener Esmeraldo-like	cromossomo <i>In silico</i> - TcChr26-S	δ-ama40/50
Tc00.1047053511903.50	CL Brener Non-Esmeraldo-like	cromossomo <i>In silico</i> - TcChr26-P	δ-ama40/50
Esm_scf7180000308152b	Esmeraldo	Scaffold - scf7180000308152 (144,278 bp)	β1-amastina
Esm_scf7180000308152a	Esmeraldo	Scaffold - scf7180000308152 (144,278 bp)	β2-amastina
Esm_scf7180000307809	Esmeraldo	Scaffold - scf7180000307809 (31,745 bp)	δ-amastina
Esm_scf7180000307960	Esmeraldo	Scaffold - scf7180000307960 (49,875 bp)	δ-ama40/50
Sylviocontig_527b	Sylvio X10/1	Contig - sylviocontig_527 (9,484 bp)	β1-amastina
Sylviocontig_527a	Sylvio X10/1	Contig - sylviocontig_527 (9,484 bp)	β2-amastina
Sylviocontig_7286	Sylvio X10/1	Contig - sylviocontig_7286 (959 bp)	δ-amastina
Sylvio_IR	Sylvio X10/1	Montagem de leituras individuais (este trabalho)	δ-ama40/50

(B)

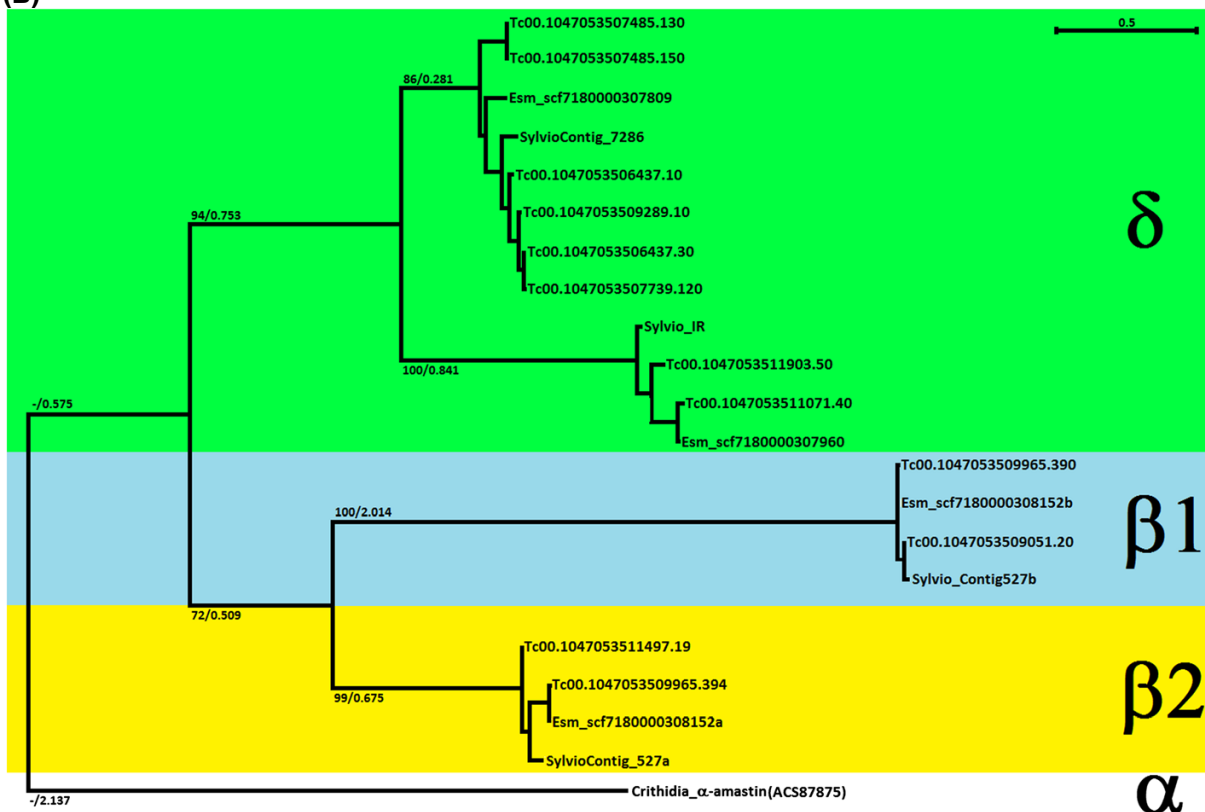


FIGURA 9 – ANÁLISE FILOGENÉTICA DAS SEQUÊNCIAS DE AMASTINAS DE DIFERENTES CEPAS DE *T. cruzi*

FONTE: Adaptado de KANGUSSU-MARCOLINO *et al.*, 2013

(A) Dados das sequências de amastinas utilizadas para gerar a árvore filogenética. (B) Árvore filogenética gerada a partir de sequência de aminoácidos de amastinas, uma sequência de α-amastina de *Angomonas* sp foi utilizada como raiz. Valores de *bootstrap* seguidos do comprimento do ramo estão mostrados nos nós basais principais.

(A)

```

δ-amastina  ----- --MSKLGAIL YGAVGFLAFL FVLVGTPIDQ FRAK---EKG ATGNTPCMTL
δ-Ama40     ----- --MSMVGAIL YGVVSFTALL LMLVGTPLDQ FRFR---TNP LPLIDPCVTL
β1-amastina MSKKKNFFVR EYGRHKGATG LLLACMVSFI FVVGTPIAM LQPK----- -LSSGTCYTL
β2-amastina -----MGFE TLAGRVGPFA YMVACISFV FATVSTPTSQ FRGRGYLDGG KESKLSGVTA

δ-amastina  WGIKEDCHST KYEFTVGEDF RECP--SVLR LFRMAEAFSI ISILLLLAAT ALGVAHFCL
δ-Ama40     WGGKLFCSGT EYIISGEDVI HEY--AVLG LFRVAEASSI VSIALLLVAT ILGVLSHLYK
β1-amastina WGTFR-PCNSP NYNWRVNWDT CKAR----RL RFQFGEAFSI CALYFAVVA- GIGSWYVLSG
β2-amastina WGVKNECTSN KYDLRSSEIK CEGNDRRLHQ LFQTTQAFSI ISIFLTFAS- IITVGMFLNG

δ-amastina  KSLKIFATLL LVVSIVTVGL VWIPMAYFYN -HDVDNCLGT PLKTCCKYGG GFVIIVIGWC
δ-Ama40     KSLRTLNLVL LVSGMTVGV VWGLMVHVYK -EKIGDCIVT PFKTQLKFGH GFGLIVASWS
β1-amastina SNKKWLTVLA SAISTVSGLV TWSMVASIHN VKLQSDTYT SAN--TKYGP GFALFVTGFV
β2-amastina KSTRRLTLLL AVASVAALLI PWACMAAVYK GSFQGTVFHD NWRDWKYSS AFGFLFGLNL

δ-amastina  LIFVAVVLLR ML----- --
δ-Ama40     LTFFAVVFLN LPYGSRLQHS CDVGDQQMHV VR
β1-amastina VQFCGLLGLI VLE----- --
β2-amastina VQLVGTVFLL AL----- --

```

(B)

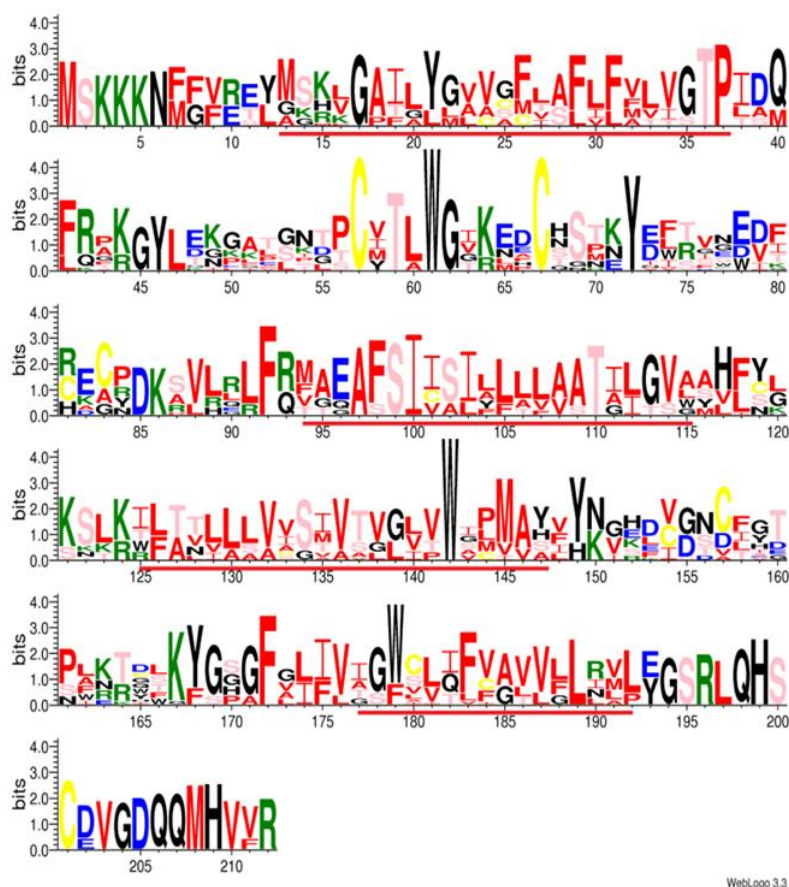


FIGURA 10 – CONSERVAÇÃO DA SEQUÊNCIA DE AMINOÁCIDOS DAS SUBFAMÍLIAS DE AMASTINAS

FONTE: Adaptado de KANGUSSU-MARCOLINO *et al.*, 2013

NOTA: (A) Sequência de aminoácidos de um representante de cada uma das subfamílias de amastina presentes no genoma de *T. cruzi*, clone CL Brener. (B) Esquema representativo da conservação dos aminoácidos das amastinas de *T. cruzi*. As prováveis regiões transmembrana estão sublinhadas em vermelho. Os aminoácidos estão representados em diferentes cores.

A variabilidade das proteínas desta família encontradas em *T. cruzi* é bastante significativa, apesar de menor que a encontrada para *Leishmania* spp. Por exemplo, em *L. major*, são encontrados 54 genes anotados como amastinas, espalhados em sete cromossomos (TritrypDB, 2014). Em uma análise *in silico* escolhendo alguns genes presentes em vários cromossomos, o tamanho das proteínas variou de 172 até 547 aa. Na mesma análise, foi avaliada a similaridade dos genes de amastinas contra um dos genes, identificado como LinJ34.0840. Os valores de similaridade encontrados apresentam alta distribuição, variando entre 25 e 100% (ROCHETTE *et al.*, 2005). Na classificação das amastinas em subfamílias, o maior ramo corresponde a δ -amastinas e está expandido em *Leishmania* spp.. Jackson (2010) sugere que essa expansão possa ser proveniente de eventos de translocação, provavelmente em resposta a uma adaptação do parasito em um hospedeiro diferente. Em uma avaliação com genes de amastinas de diferentes espécies de *Leishmania*, foram encontrados grupos de genes exclusivos para determinadas espécies, o que pode ser consequência da pressão do ambiente ou da especiação (REAL *et al.*, 2013).

Dos seis genes de amastinas encontrados para cada alelo no genoma de CL Brener, três são de δ -amastinas, organizados em sequência e intercalados pelo gene de tuzina no cromossomo 34; um corresponde a δ -Ama40, que está acompanhado por um pseudogene de tuzina, no cromossomo 26; e os outros dois são de β -amastinas, sendo um de β 1- e outro de β 2-amastina, organizados em sequência, com uma grande região espaçadora entre os dois genes, no cromossomo 32 (WEATHERLY *et al.*, 2009).

Esta organização genômica foi confirmada para os genes de β 2-amastina, e δ -Ama40 em ensaios de PFGE seguido de hibridação com sondas específicas. A figura 11 (a e b) mostra que foi marcada apenas uma banda cromossômica para cada hibridização com as sondas específicas de β 2-amastina, e δ -Ama40, apresentando os tamanhos aproximados dos cromossomos 32 e 26, que têm 968 e 834 Mpb respectivamente (WEATHERLY *et al.*, 2009). Para a hibridação com δ -amastina (figura 11c), evidenciaram-se três bandas cromossômicas de tamanhos

diferentes para a cepa G e os clones CL Brener e Dm28c, e uma banda cromossômica apenas para as cepas Colombiana, Sylvio X-10 e Y. O mesmo padrão foi encontrado na hibridização com a sonda para o gene de tuzina (figura 11d), mostrando que estes realmente acompanham os genes de δ -amastinas e não as demais isoformas.

É bastante comum a variação no tamanho e número de bandas cromossômicas de diversas cepas de *T. cruzi* quando alguns genes são utilizados como sonda. Por exemplo, os genes para delta-6 desnaturase, transportador de hexose e proteína hipotética TEUF0180 são marcadores para a banda cromossômica XX e foram assinalados no genoma de CL Brener no cromossomo TcChr37 em ambos os alelos (Esmeraldo e não-Esmeraldo). Quando usados como sonda em *Southern blot* de corridas de PFGE, marcaram bandas com tamanho correspondente à banda cromossômica XX (3,27 Mb) para as cepas Y, clone Esmeraldo-cl3 do grupo TcII; para os clones SO3-cl5 do grupo TcV; e para o clone CL Brener do grupo TcVI. Para as cepas do grupo TcI a hibridização ocorreu em bandas de tamanhos menores, variando entre as cepas testadas (clone Dm28c, cepa G e isolado Tc1161). Para vários marcadores foi encontrado o mesmo padrão de bandas nas diferentes cepas. Assim como na análise similar, feita com marcadores para o cromossomo TcChr7. Foi possível observar a hibridização em duas bandas (0,96 e 1,02 Mb) para as cepas de grupo TcI, exceto cepa G, que apresentou apenas uma banda (0,96 Mb). Para as cepas dos demais grupos foi encontrada marcação para uma banda (0,77 Mb), exceto para Esmeraldo-cl3, que mostrou duas bandas (0,87 e 0,98 Mb). Estes dados mostram que há forte sintonia entre as diferentes cepas de *T. cruzi*, contudo há divergências entre a montagem do genoma e os cariótipos diferem entre as cepas, assim como foi mostrado em nossas análises (figura 11) (SOUZA *et al.*, 2011). Na avaliação de SOUZA *et al.*, 2011, isto pode dever-se a não disjunção de cromossomos durante a mitose, pela presença de cromossomos homólogos de tamanhos diferentes ou a irregularidades nas trocas genéticas destes organismos.

Frente às possibilidades de erros na montagem do genoma, principalmente em se tratando de famílias multigênicas, e visando obter uma avaliação mais acurada da organização genômica dos genes de amastinas, foi realizada uma análise do número de leituras individuais obtidas no sequenciamento do genoma de CL Brener. A fim de quantificar/estimar o número de cópias de cada um dos genes de amastinas e comparar com os dados de *Southern blot* dos ensaios de PFGE, o Dr Rondon M. de Pessoa Neto (UFMG) realizou uma análise de bioinformática utilizando as leituras individuais obtidas no sequenciamento do genoma de CL Brener (13x a cobertura do genoma). Esta análise mostrou que os genes das δ -amastinas estão presentes em um número mais elevado de cópias comparado a montagem do genoma. Estimou-se que existam 14 cópias dos genes de amastinas no genoma de CL Brener, sendo 2 β -amastinas e 12 δ -amastinas. Estas evidências sugerem que as 12 cópias encontradas podem estar espalhadas em outros cromossomos detectados nos ensaios de *Southern blot*, e/ou em maior número de cópias por cromossomo. Em *Leishmania*, as δ -amastinas estão presentes em quatro cromossomos diferentes, e em loci diferentes no cromossomo 34 (JACKSON, 2010).

Estes genes de δ -amastinas são aparentemente muito transponíveis, e podem ter se translocado no genoma (JACKSON, 2010). Sendo assim, os dados de PFGE em conjunto com as análises de leituras individuais mostram que o mesmo evento pode ter ocorrido em algumas cepas de *T. cruzi* com os genes destas proteínas. Para este parasito é bastante conhecida a presença de elementos retrotransponíveis, como SIRE, VIPER, DIRE, L1Tc, NARTc em regiões vicinais às MASPs (BARTHOLOMEU *et al.*, 2009), ou espalhados no genoma (BRINGAUD *et al.*, 2002; PAVIA *et al.*, 2012; SOUZA *et al.*, 2011). Em *Leishmania* sp., um elemento transponível do tipo SIDER1 foi identificado na região 3'UTR de diversos genes, incluindo amastinas. SIDER1 parece regular a expressão de mRNA de modo estágio-específico, não somente para aumentar a expressão, o que sugere que é necessária a atuação de elementos regulatórios adicionais para a regulação da expressão de mRNA em *Leishmania* sp. (ROCHETTE *et al.*, 2008).

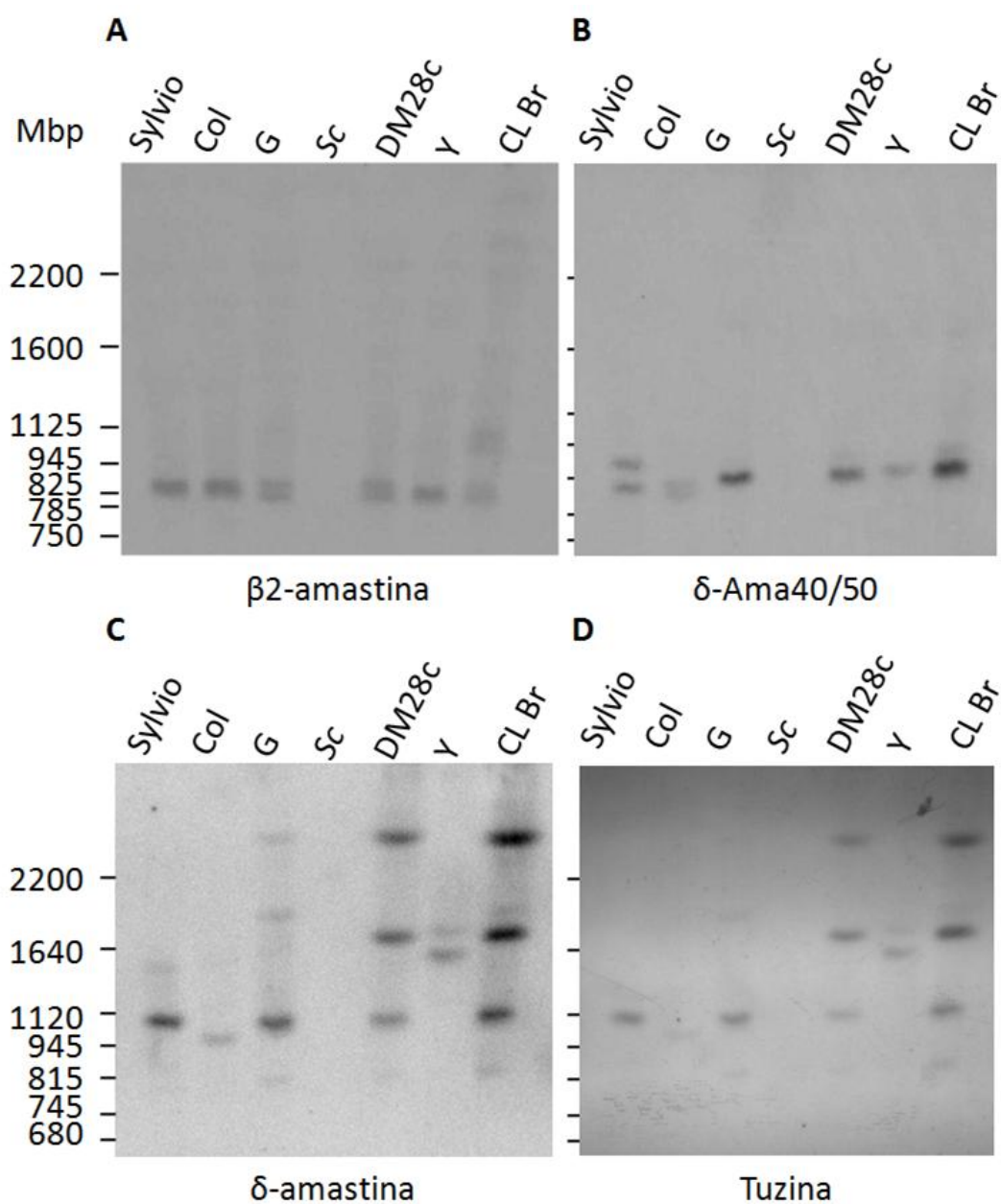


FIGURA 11 – LOCALIZAÇÃO GENÔMICA DOS GENES DE AMASTINAS PARA DIFERENTES CEPAS/CLONES DE *T. cruzi*

FONTE: Adaptado de KANGUSSU-MARCOLINO *et al.*, 2013

NOTA: Bandas cromossômicas de diferentes cepas/clones de *T. cruzi* foram separadas por eletroforese de campo pulsado (PFGE) e transferidos para membranas e hibridizados com sondas para o genes $\beta 2$ -amastina (A), δ -Ama40 (B), δ -amastina (C) e tuzina (D) marcadas com ^{32}P -dCTP. As cepas/clones utilizados foram SylvioX-10 (Sylvio), Colombiana (Col.), G and Dm28c, Y and CL Brener (CLBR). Os tamanhos das bandas cromossômicas de levedura (SC estão indicados a esquerda.

Para os genes de $\beta 1$ - e $\beta 2$ -amastinas, o número de cópias parece corresponder ao obtido na montagem do genoma de CL Brener (JACKSON, 2010; WEATHERLY *et al.*, 2009). A presença de dois genes de β amastinas é observada também para *Leishmania seymouri*, mas é diferente do encontrado em *Angomonas* sp. e *L. major*, que apresentam três genes arranjados em sequência. Contudo, os genes vizinhos são sintênicos nos organismos citados por Jackson (2010).

6.1.2 Expressão de mRNA das subfamílias de amastinas de *T. cruzi*

A avaliação das sequências de DNA das amastinas de *T. cruzi* revelar alta divergência, propondo e até uma nova subdivisão destes genes. Enquanto as análises de PFGE mostraram uma grande variação no número de cópias e no perfil de hibridação entre cepas. Por isso, é necessário avaliar se a expressão destas proteínas ocorre de forma semelhante ou distinta da descrita para δ -amastina (TEIXEIRA *et al.*, 1994). Por análises de *northern blot* foi utilizando *T. cruzi* (clone Sylvio X-10, cepa Colombiana, cepa G, clone Dm28c, cepa Y, clone CL Brener) nas formas epimastigota, tripomastigota e amastigota. Detectamos que tanto $\beta 1$ - quanto $\beta 2$ -amastina são expressas, e em níveis mais elevados em formas epimastigotas, podendo apresentar alguma função biológica diferente das outras amastinas, que são mais expressas em amastigotas (figura 12). Cabe ressaltar que este resultado está de acordo com o descrito para a cepa Brazil, cujos dados de microarranjos revelaram que, diferentemente das amastinas descritas por Teixeira *et al.*, (1994), as β -amastinas são mais expressas em forma epimastigota (MINNING *et al.*, 2009).

Nesses experimentos de *northern blot* foram detectados no clone CL Brener 2 transcritos de $\beta 2$ -amastina com tamanhos distintos, o que indica um possível processamento alternativo do mRNA ou divergência entre os alelos. A presença de transcritos de tamanhos diferentes já foi encontrada por RT-PCR anteriormente para o gene de LYT1, uma proteína envolvida na hemólise com importante papel na infectividade. Este evento ocorre por *trans-splicing* alternativo, o transcrito de maior tamanho codifica para a proteína completa e está expresso somente em

tripomastigota (MANNING-CELA *et al.*, 2002). Posteriormente foi descrito que estes transcritos geram proteínas com localização e atividade diferentes (BENABDELLAH *et al.*, 2007).

Por captura de *spliced leader*, uma abordagem em larga escala utilizando mRNA-SEQ, foi mostrado que *T. brucei* pode processar o mRNA de formas alternativas. Foram encontrados 2500 eventos de *splicing* alternativo, o que pode alterar a região regulatória 5'UTR ou o códon de início de tradução, alterando drasticamente a proteína ou a regulação da expressão do gene (NILSSON *et al.*, 2010). Assim, é interessante obter mais informações sobre estes transcritos de tamanho diferentes observados para β 2-amastina. Como não foi detectado em outras fases do ciclo, pretendemos avaliar a estabilidade ao longo do crescimento de formas epimastigotas. Podemos ainda mapear o sítio de adição e de *splice leader* da cauda poli-A.

Para as δ -amastinas, tanto as clássicas como as δ -Ama40, a expressão de mRNA foi maior em amastigota, exceto para cepa G, que expressa δ -Ama40 em níveis mais baixos que para outras cepas e δ -amastina em níveis praticamente não detectáveis. Comparando as bandas cromossômicas identificadas pela hibridização da sonda de δ -amastina (figura 11c), pode-se observar para a cepa G que a proporção da intensidade entre as duas bandas mais pesadas e a banda mais leve é diferente da proporção obtida para os clones Dm28c e CL Brener. Sendo assim a cepa G parece apresentar menos cópias do gene de δ -amastina no genoma que os clones Dm28c e CL Brener. Contudo, provavelmente esta não é uma explicação para a menor expressão de δ -amastina em cepa G, já que o clone Sylvio X-10 e a cepa Y apresentam apenas uma banda cromossômica e a expressão de δ -amastina para estas cepas é comparável a encontrada em CL Brener.

A expressão diferenciada de epítomos na superfície de diferentes cepas, principalmente na forma biológica presente no hospedeiro, já foi amplamente relatada e pode ser utilizada como diagnóstico para identificação de cepas que o paciente tem em maior abundância, podendo direcionar o tratamento e traçar o perfil de cepas responsáveis pela doença, sendo mais ou menos imunogênicas (MENDES

et al., 2013). Uma comparação de reatividade de formas amastigotas extracelulares (AEs) de diferentes cepas de *T. cruzi* foi realizada tratando as culturas com anticorpos contra amastigotas. Foi observado que a cepa G, cujas formas AEs são consideradas mais infectivas para célula HeLa, apresentou maior reatividade ao anticorpo 1D9 que diversas outras cepas testadas, entre elas a cepa Y e os clones CL Brener e Sylvio X-10 (VERBISCK *et al.*, 1998). Em vista da baixa expressão de δ -amastina em amastigota de cepa G, Cruz *et al.* (2012) propõe que esta proteína seja um regulador negativo da infectividade de formas amastigotas extracelulares.

Por dados de microarranjo, foi mostrado que a cepa Brazil, pertencente ao grupo TcI, apresenta elevação da expressão de δ -Ama40 em amastigota, quando comparado com tripomastigota, epimastigota e tripomastigota metacíclico (MINNING *et al.*, 2009). Comparando este dado com os obtidos neste trabalho para cepas do grupo TcI, o perfil de expressão do clone Sylvio X-10, mostrado somente em epimastigota e amastigota, são mais próximos aos encontrados anteriormente que o perfil da cepa G. Este perfil não é exatamente igual ao encontrado para a δ -amastina pois a δ -Ama40 parece ter maior expressão tanto em tripomastigota como em epimastigota. Mais experimentos são necessários para relacionar estes dados com algum fenótipo.

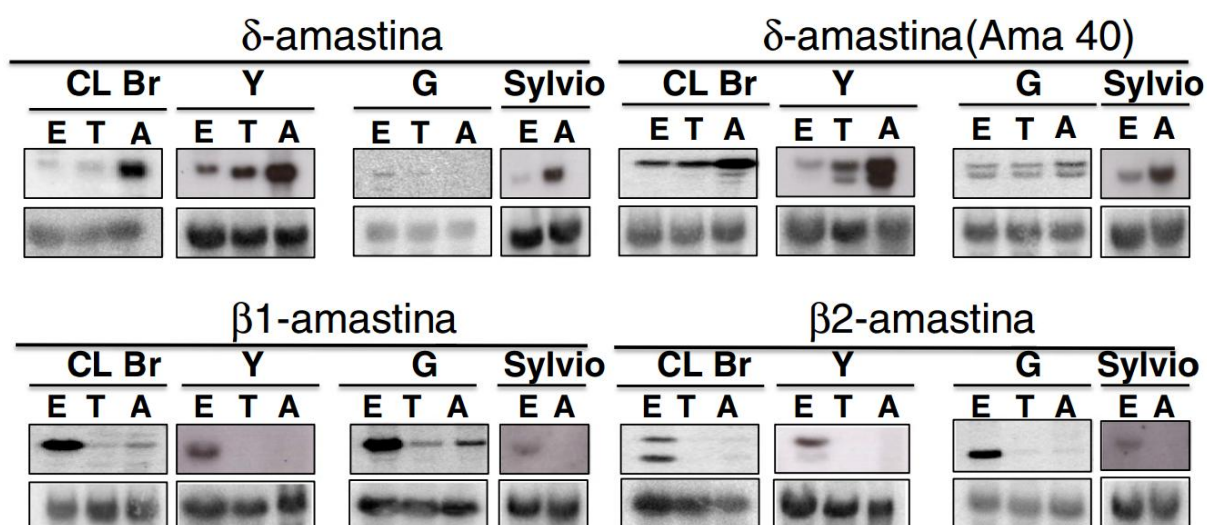


FIGURA 12 – EXPRESSÃO DO mRNA DE AMASTINAS DURANTE O CICLO BIOLÓGICO DE *T. cruzi* EM DIFERENTES CEPAS/CLONES

FONTE: Adaptado de KANGUSSU-MARCOLINO et al., 2013

NOTA: RNA total foi extraído de formas epimastigotas (E), tripomastigotas derivada de cultura (T) e amastigotas (A) das cepas/clones CL Brener, Y, G e Sylvio X-10. Estas amostras foram submetidas a eletroforese (~10 µg/ canaleta) e transferidas para membranas de nylon, que foram hibridizadas com sondas marcadas com α -32P-dCTP para os genes de δ -amastina, δ -Ama40, β 1- and β 2-amastinas. Os painéis inferior mostram o controle de aplicação, para o qual as mesmas membranas foram hibridizadas com sonda de um fragmento de rRNA 24S.

Em *Leishmania* a avaliação por *northern blot* da expressão de 18 genes de amastinas mostrou que a maioria delas é mais expressa em forma amastigota, com exceção do gene LinJ34.0730, cuja maior expressão está em forma promastigota em fase estacionária (ROCHETTE et al., 2005). Por ensaios de *microarray* foi mostrado que os genes anotados como amastinas apresentam um aumento na expressão variando entre 1,9 até 13,9 vezes em formas amastigotas (ROCHETTE et al., 2009).

6.1.3 Localização subcelular das amastinas de *T. cruzi*

Além da expressão diferencial durante o ciclo biológico, outra característica não descrita para todas as subfamílias de amastinas é a localização subcelular. São observadas, a partir das análises de sequências, as topologias das quatro

subfamílias de amastinas em *T. cruzi* (figura 13). Na predição realizada com uso do software Sosui (HIROKAWA *et al.*, 1998), todas elas apresentam quatro regiões transmembrana e os aminoácidos que fazem parte da assinatura das amastinas, porém apresentam algumas diferenças nas regiões hidrofílicas. A δ -Ama40 apresenta uma porção hidrofílica C-terminal 19 aa mais longa que as outras, enquanto as β 1- e β 2-amastinas apresentam uma porção hidrofílica N-terminal mais longa com 11 e 5 aa, respectivamente, sendo que em β 1 é mais longa que β 2. Mesmo que todas apresentem topologia semelhante, contendo quatro regiões transmembrana, as diferenças citadas podem trazer consequências para a localização das proteínas. Diante destas variações observadas, é interessante analisar a localização subcelular.

Para tanto, uma representante de cada família foi expressa em fusão com GFP em formas epimastigotas de *T. cruzi* da cepa G ou do clone CL Brener utilizando como base o vetor pTREXnGFP (DAROCHA *et al.*, 2004b). Após obtenção de formas epimastigotas expressando de forma estável, foram obtidas imagens por microscopia confocal. As imagens obtidas demonstraram que, enquanto GFP sozinha está presente em todo o citoplasma do parasito (figura 14A-C). Fusões de GFP com β 1- e δ -amastinas são claramente localizadas na superfície da célula (figura 14 D-F e J-L). Curiosamente, foi observada uma localização celular diferente para δ -Ama40::GFP, com um padrão pontuado no citoplasma do parasito, bem como uma distribuição aparentemente dispersa dentro do citoplasma para a β 2-amastina::GFP (figura 14 G -I e MO). A principal diferença topológica entre a δ -Ama40 e as outras amastinas é a cauda C-terminal hidrofílica, podendo esta ser responsável pela diferença na localização desta proteína. Esta possibilidade precisa ser testada.

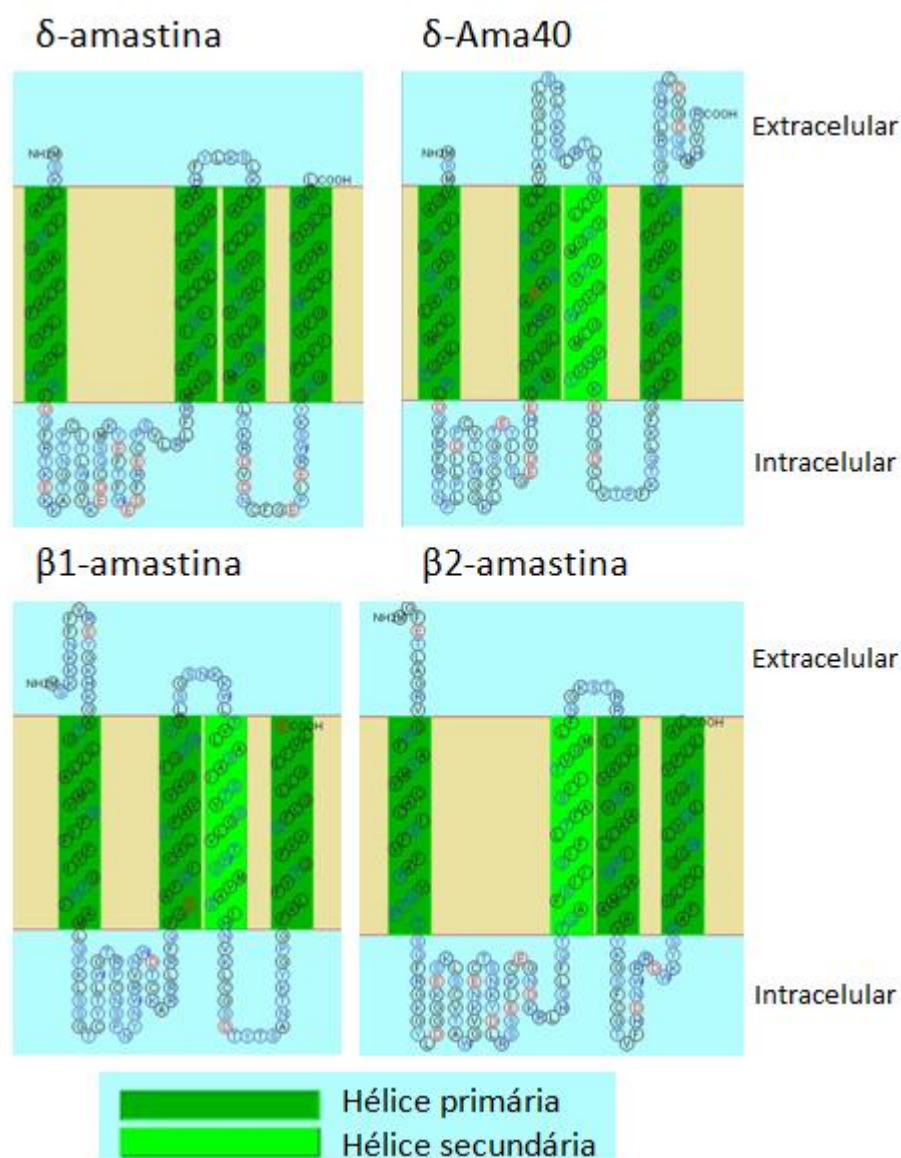


FIGURA 13 – PREDIÇÃO DA TOPOLOGIA DAS SUBFAMÍLIAS DE AMASTINAS

FONTE: O Autor (2014)

NOTA: A predição topológica (SOSUI, HIROKAWA *et al.*, 1998) foi feita com a sequência de aminoácidos para os seguintes genes: TcCLB.507485.130, TcCLB.511071.40, TcCLB.509965.390 e TcCLB.509965.394, das respectivas famílias: δ -, δ -Ama40, β 1 e β 2-amastina.

Como mencionado acima, δ -Ama40::GFP apresentou uma localização pontual ao longo do citoplasma, provavelmente associada a membrana de vesículas intracelulares. A fim de caracterizar a provável organela identificada, foram realizados ensaios de co-localização utilizando como marcador anticorpos que

reconhecem a enzima fosfoenolpiruvato carboxiquinase (PEPCK), uma enzima glicossomal (LINSS *et al.*, 1993) e TcRBP40, uma proteína localizada em reservossomos (GUERRA-SLOMPO *et al.*, 2012). As imagens confocais de imunofluorescência (figura 15) utilizando anti-PEPCK mostram que não há co-localização das proteínas δ -Ama40::GFP com uma proteína glicossomal. Portanto, esta amastina parece não estar localizada em glicossomos. Por outro lado os anticorpos anti-TcRBP40 não funcionaram em nossos ensaios. Novos experimentos como fracionamento de organelas e microscopia eletrônica são necessários para melhor definir a localização desta amastina.

Um ponto normalmente levantado na caracterização de proteínas é se a fusão com uma proteína repórter alteraria a distribuição celular em relação a proteína nativa. No caso da localização das δ -amastinas, a localização em membrana plasmática foi confirmada com estudos anteriores, quando foram localizadas na membrana de formas amastigotas de *T. cruzi* por imunolocalização utilizando antisoro gerado contra peptídeos de amastinas (TEIXEIRA *et al.*, 1994). Para confirmar a localização das outras famílias de amastinas seria necessária a realização de outros experimentos utilizando anticorpos específicos ou utilizando outras *tags*. Especialmente no caso da δ -Ama40, já que há suspeita da cauda C-terminal influenciar na localização da proteína. A produção de anticorpos específicos para as diferentes amastinas está em andamento em nosso laboratório em parceria com o laboratório da Prof.^a Santuza Teixeira, na UFMG.

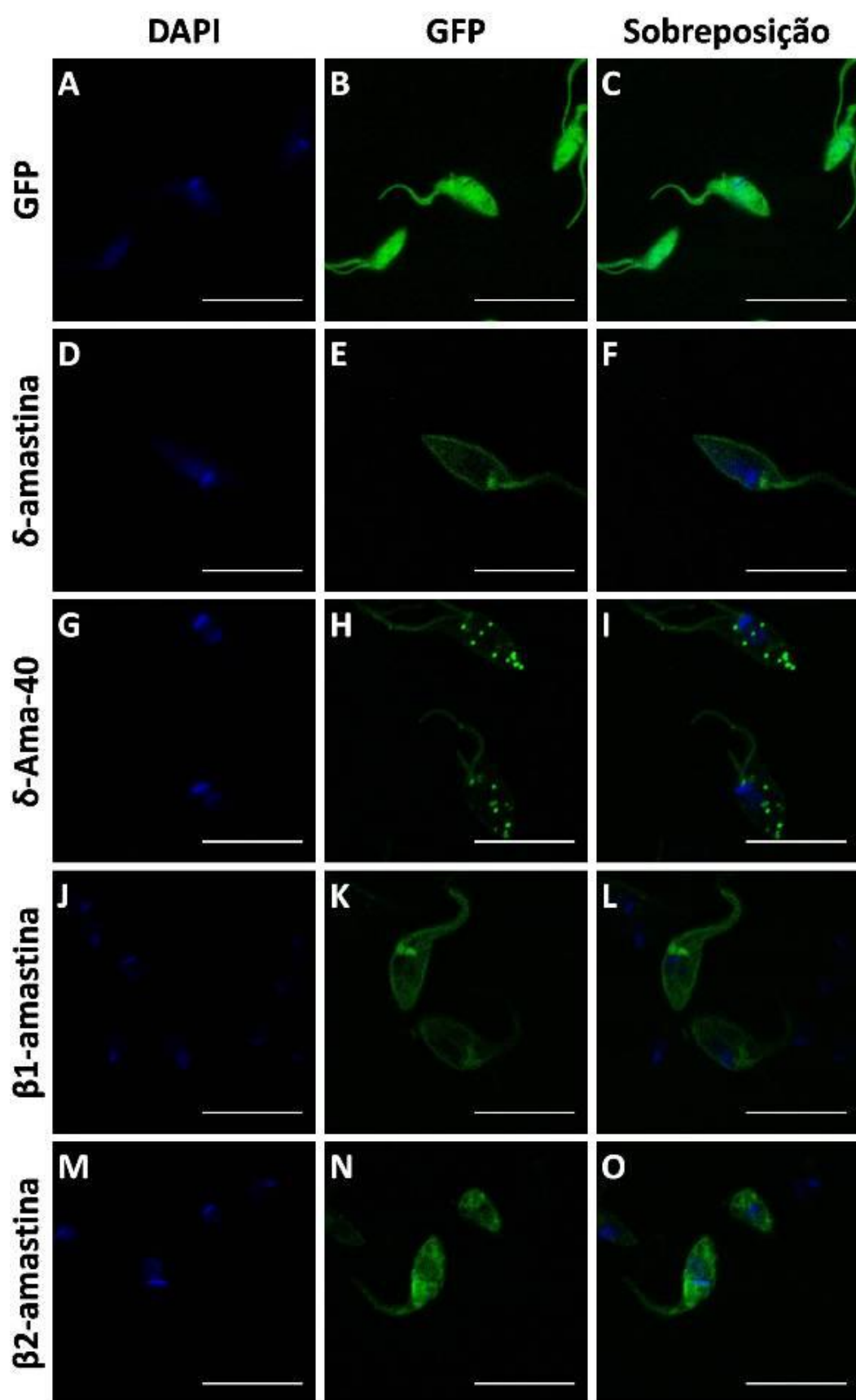


FIGURA 14 – LOCALIZAÇÃO SUBCELULAR DE DIFERENTES AMASTINAS EM FUSÃO COM GFP.

FONTE: Adaptado de KANGUSSU-MARCOLINO et al., 2013

NOTA: Imagens de populações estáveis de epimastigotas do clone CL Brener ou cepa G obtidas por microscopia confocal usando 1000X de aumento e 2,2 de zoom digital. Os painéis A-C mostram parasitos transfectados com um vetor para expressão somente de GFP. Os parasitos mostrados nos painéis seguintes foram transfectados com vetores para expressão das diferentes amastinas em fusão com GFP, sendo (D-F) corresponde a δ -amastina, (G-I) a δ -Ama40, (J-L) a β 1-amastina e (M-O) a β 2-amastina. Coloração com DAPI está mostrada nos painéis (A, D, G, J e M), a fluorescência de GFP pode ser vista nos painéis (B, E, H, K e N), e nos painéis C, F, I, L e O) encontram-se as imagens sobrepostas. Barra = 10 μ m.

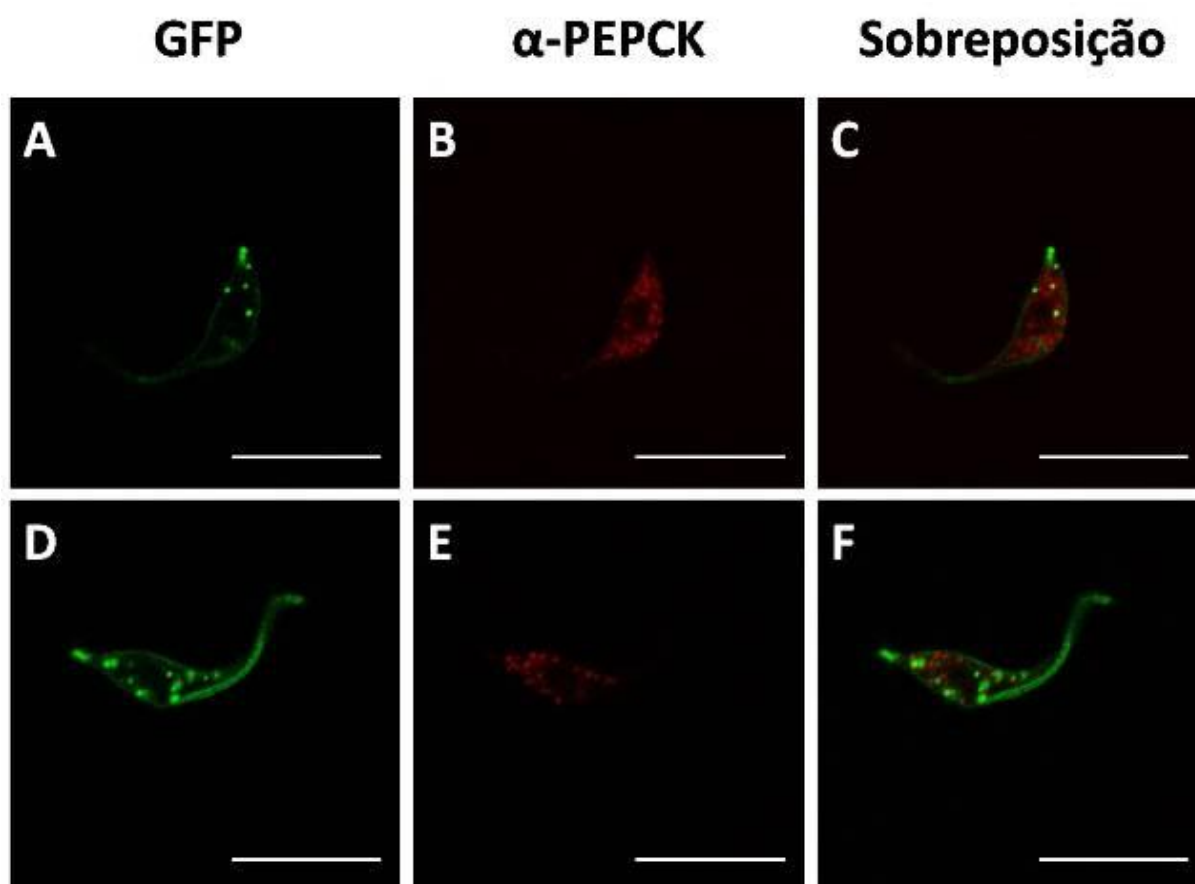


FIGURA 15 – LOCALIZAÇÃO SUBCELULAR DE δ -AMA-40 EM FUSÃO COM GFP

FONTE: Adaptado de KANGUSSU-MARCOLINO et al., 2013

NOTA: Populações estáveis de epimastigotas de CL Brener, transfectadas com pTREXAma40GFPNeo foram incubadas com o anticorpo α -PEPCK e com o anticorpo secundário conjugado com Alexa546. Os painéis mostram imagens obtidas por microscopia confocal usando 1000X de aumento e 2,2 de zoom digital. A fluorescência de GFP está mostrada nos painéis A e D e a fluorescência de Alexa 546 está nos painéis B e E. Nos painéis C e F estão as sobreposições das imagens. Barra = 10 μ m.

A partir das análises por microscopia confocal, foi observado que o padrão de localização das subfamílias de amastinas fusionadas a GFP está relacionado a determinadas estruturas no parasito, ou a superfície ou estruturas citoplasmáticas. A fim de identificar se estas proteínas estão associadas a membranas ou não, foi realizada a separação por ultracentrifugação das frações solúvel e insolúvel do parasito. A figura 16 mostra a avaliação destas frações e das proteínas totais por ensaios de *western blot* utilizando o anticorpo α -GFP. Todas as amastinas expressas em populações resistentes a G418 foram detectadas na fração insolúvel (I), o que indica que todas as amastinas possam estar associadas a membranas, diferentemente da proteína verde fluorescente (GFP), que, como esperado, foi detectada na fração solúvel. Estes dados estão de acordo com a topologia encontrada para os quatro grupos de amastinas testados, que mostrou a presença de domínios transmembranas (figura 13).

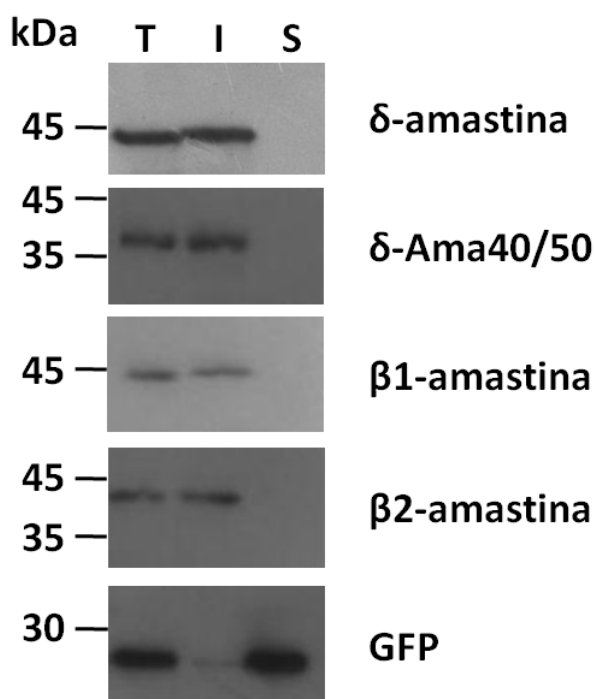


FIGURA 16 – DISTRIBUIÇÃO DAS PROTEÍNAS AMASTINAS EM FRAÇÕES CELULARES

FONTE: O autor (2014)

NOTA: Imunoblot das frações total (T), de Insolúvel (I) e Solúvel (S) de epimastigotas expressando δ -Ama, δ -Ama40, β 1- e β 2-amastinas em fusão com GFP. Todas as membranas foram incubadas com anticorpo α -GFP.

Os dados obtidos até aqui foram estruturados e publicados na revista BMC Microbiology (KANGUSSU-MARCOLINO *et al.*, 2013– ver anexo 2).

Das amastinas expressas em fusão com GFP, as duas β -amastinas são mais expressas, no parasito selvagem, em epimastigotas e as duas δ -amastinas em amastigotas. A “superexpressão” de proteínas endógenas pode alterar o fenótipo dos parasitos de alguma forma, como constatado previamente por outros trabalhos (CRUZ *et al.*, 2012; PEREIRA *et al.*, 2003; SCHOIJET *et al.*, 2011). Para iniciar a avaliação destes efeitos, foram realizadas etapas de clonagem de parasitos a partir das populações resistentes a G418 transfectadas com cada uma das construções, a fim de obter maior homogeneidade da cultura, uma vez que a fluorescência de algumas populações resistentes estava abaixo de 50%. Após a clonagem, todas as culturas tiveram mais que 90% de fluorescência, exceto para β 2-amastina em Cepa G, que mesmo após várias tentativas de clonagem, todos os clones obtidos tiveram no máximo 10% de fluorescência em fase log.

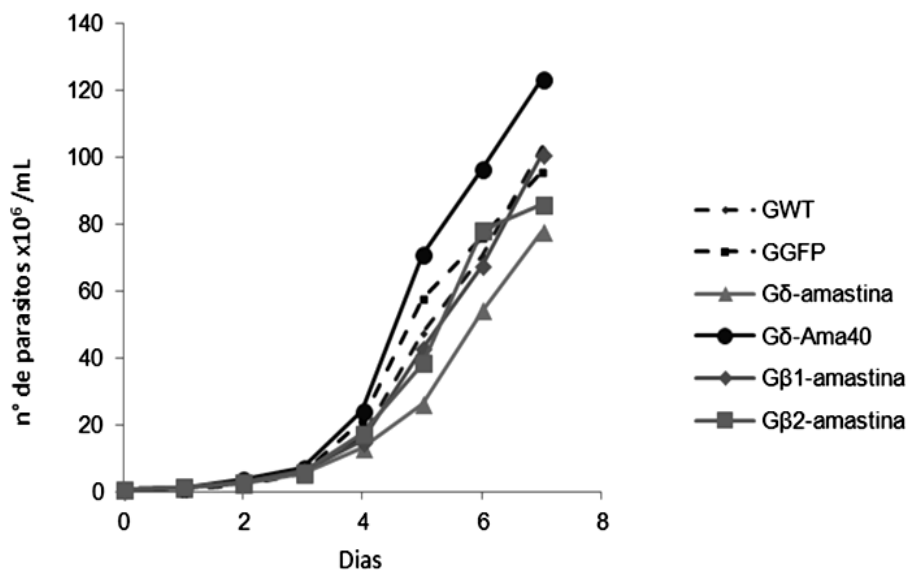
Após a inserção de vetores para expressão das amastinas de diferentes subfamílias fusionadas a GFP em formas epimastigotas da cepa G e clone Sylvio-X10 (ambas do grupo TcI), foram obtidos clones por diluição seriada, que foram testados em relação a sua taxa de crescimento. Na figura 17 temos dados preliminares da curva de crescimento obtida de transfectantes da cepa G (A) e do clone Sylvio X-10 (B), resultantes de um único experimento.

Preliminarmente, os gráficos apresentados na figura 17, mostram que a expressão de determinadas amastinas (p.ex.: δ -amastina) tem um efeito mais drástico no crescimento de parasitos gerados a partir do clone de Sylvio X-10, quando comparado aos clones derivados da cepa G. Nesse experimento, o perfil de fluorescência das culturas foi avaliado por citometria de fluxo, a fim de avaliar o grau de variação na expressão das proteínas de fusão frente ao envelhecimento das culturas (figura 18). Os gráficos, expressos em número de eventos por intensidade de fluorescência captada por FL1 (530 ± 30 nm) mostram os espectros de fluorescência obtidos para as culturas selvagens e clones dos transfectados (indicados no eixo Y) no decorrer de 12 dias após o inóculo em início de fase log (5 x

10^5 parasitos/ mL). A intensidade de fluorescência emitida pelos parasitos selvagens foi considerada 0%. Os eventos que apresentaram maior intensidade de fluorescência, dentro da faixa indicada na porção superior dos gráficos, foram usados para calcular a porcentagem fluorescente da população, indicados nas tabelas na porção inferior da figura 18. Avaliando os resultados obtidos, é possível observar um deslocamento dos espectros, indicando diminuição da fluorescência das populações a medida que a cultura envelhece. A confirmação deste resultado pode indicar que a expressão regulada pelas reigões intergênicas de GAPDH (vetor pTREX) diminui em parasitos metacíclicos.

Apesar destas alterações detectadas, estes experimentos precisam ser realizados em triplicata com diferentes clones derivados da mesma cultura. Pode-se destacar que foi descrito para a cepa G que a expressão de δ -amastina fusionada a GFP, não afetou a taxa de crescimento (CRUZ *et al.*, 2012).

(A)



(B)

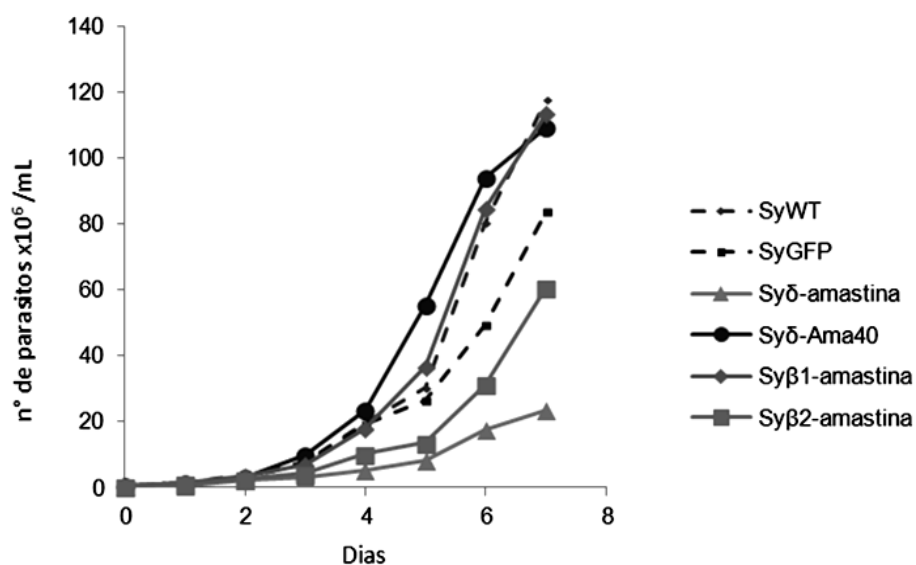
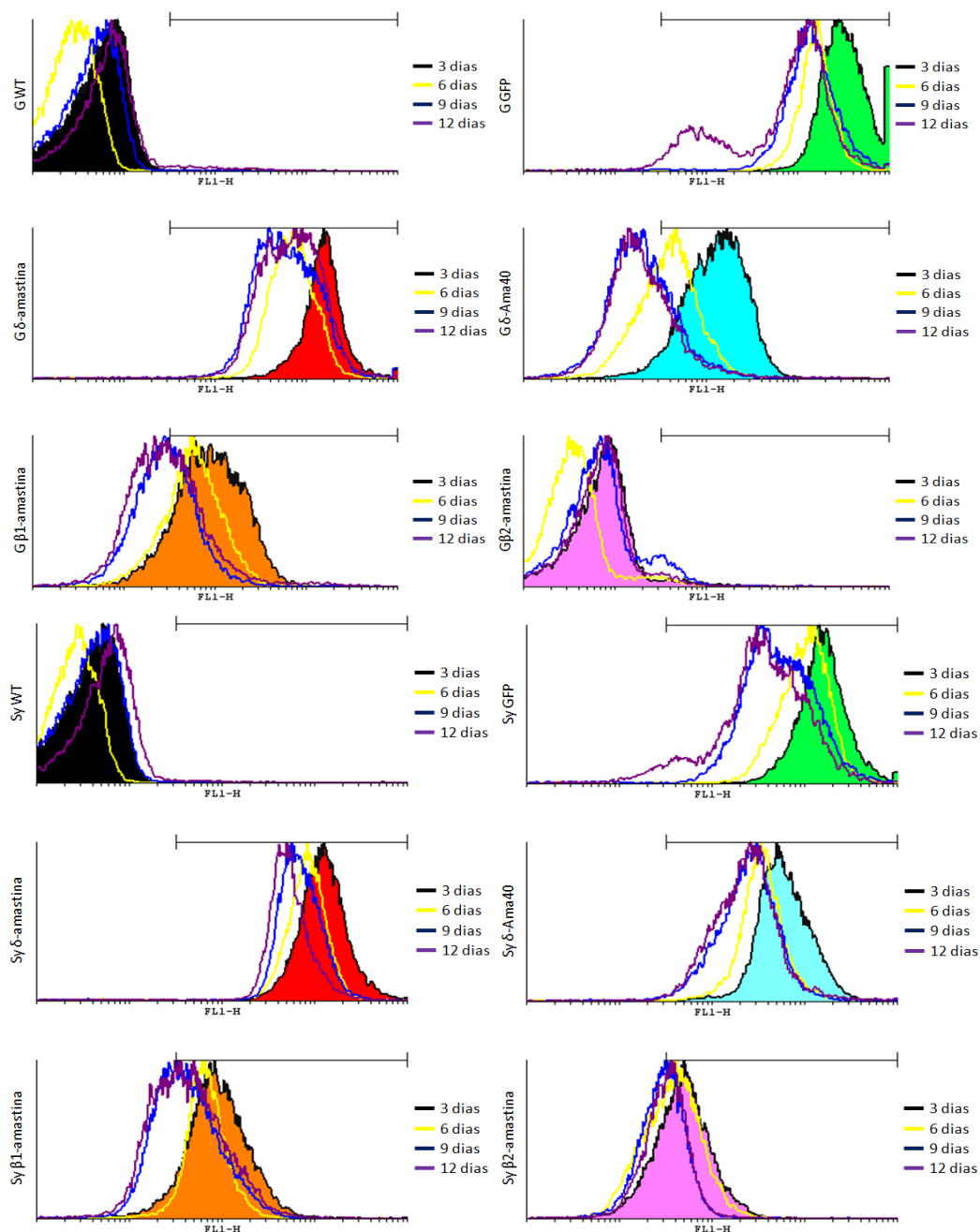


FIGURA 17 – EFEITO DA EXPRESSÃO DE AMASTINAS FUSIONADAS GFP NO CRESCIMENTO DE EPIMASTIGOTAS.

FONTE: O autor (2014).

NOTA: A legenda das curvas de crescimento das formas epimastigotas está indicada a direita do gráfico. GWT e SyWT correspondem a cepa G selvagem e o clone Sylvio X-10 selvagem, respectivamente. O número de parasitos por mL é representado no eixo Y e o tempo de contagem no eixo X. Os clones obtidos das populações transfectadas com pTREXnGFP, para expressão de GFP fusionada ou não com as amastinas (δ -, δ -Ama40, β 1- e β 2), e as populações selvagens (Cepa G e Sylvio X-10) foram inoculados a $0,5 \times 10^6$ parasitos/mL e o crescimento foi acompanhado diariamente por contagem em câmara de Neubauer.

(A)



(B)

PARASITOS	PORCENTAGEM DE PARASITOS FLUORESCENTES				PARASITOS	PORCENTAGEM DE PARASITOS FLUORESCENTES			
	3º Dia (%)	6º Dia (%)	9º Dia (%)	12º Dia (%)		3º Dia (%)	6º Dia (%)	9º Dia (%)	12º Dia (%)
G GFP	99,1	99,4	99,4	98,4	Sy GFP	99,5	99,5	99,1	98,8
G δ-amastina	99,3	99,6	99,5	99,4	Sy δ-amastina	99,5	99,5	99,6	99,5
G δ-Ama40	95,9	64,1	27,6	26,6	Sy δ-Ama40	99,5	99,7	99,5	99,7
G β1-amastina	89,2	80,8	46,6	46,3	Syβ1-amastina	95,1	88,9	89,9	92,4
G β2-amastina	3,5	2,5	7,9	3,8	Syβ2-amastina	98,4	99,1	92,8	90,0

FIGURA 18 – INFLUÊNCIA DO CRESCIMENTO DE FORMAS EPIMASTIGOTAS NA EXPRESSÃO DAS PROTEÍNAS DE FUSÃO A GFP

FONTE: O autor (2014).

NOTA: (A) Citometria de fluxo de formas epimastigotas em diferentes pontos da curva de crescimento. Foram retiradas alíquotas para citometria das amostras da curva de crescimento (figura 17) nos dias 3, 6, 9 e 12 após o inóculo (dia 0). O nome da amostra está indicado no eixo Y, sendo representado pela abreviação da cepa seguida do nome da proteína que expressa ou da designação WT, para os parasitos não transfectados. Sy corresponde ao clone Sylvio e G corresponde a Cepa G. No eixo Y é representado o número de evento e no eixo X a intensidade de fluorescência de GFP, adquirida no filtro FL1 que capta a frequência de 530 nm com variação de 30 nm para mais ou para menos (530/30). (B) Compilação dos dados de porcentagem de parasitos fluorescentes calculados com o software Cyflogic, usando a amostra WT como 0% de fluorescência.

6.1.4 Deleção dos genes que codificam isoformas das proteínas amastinas de *Trypanosoma cruzi*

Nos estudos de genômica funcional o bloqueio da expressão gênica por meio da geração de nocautes por recombinação homóloga, ou por RNAi são importantes etapas. Como mencionado anteriormente, diversas dificuldades são encontradas na aplicação de tais estratégias em *T. cruzi*, sendo que somente o nocaute clássico foi utilizado com sucesso para diferentes genes (COOPER *et al.*, 1993; DE SOUZA *et al.*, 2010; MACRAE *et al.*, 2006). Os genes que codificam as diferentes amastinas de *T. cruzi* estudadas neste trabalho estão presentes em cromossomos diferentes. Em virtude do elevado número de cópias estimado para as δ -amastinas, e a não acurácia da montagem do genoma, não foi possível desenhar uma estratégia de nocaute para estes genes. Assim, planejamos fazer a deleção de δ -Ama40, presente no cromossomo 26 em cópia única, e das β -amastinas, no cromossomo 32.

Para deleção da δ -Ama40, no cromossomo 26, foram desenhados iniciadores, baseados no genoma de CL Brener (TRITRYPDB, 2011), para amplificação da porção final do pseudogene de tuzina a jusante da sequência da δ -Ama40. Esta região foi chamada de UP26 e apresenta 414 pb. Foram desenhados também, iniciadores para amplificação da porção inicial do gene de uma proteína hipotética a montante da sequência da δ -Ama40, sem alteração da região codificadora e região regulatória 5' deste gene, já que foi utilizado um sítio de *Bam*HI

interno para obtenção do cassete de inserção no genoma. Esta região foi chamada de DN26, e contém 740 pb (figura 19A).

De forma semelhante, utilizando o genoma de CL Brener como referência, foram desenhados iniciadores para amplificação de regiões chamadas UP32 e DN32, com 420 e 407 pb, respectivamente. Para deleção das β -amastinas, as duas regiões de recombinação foram escolhidas de a fim de truncar os dois genes simultaneamente. Na região UP32 foi inserido um códon de parada no primer, de forma a bloquear a tradução da proteína. A região DN32 seria recombinada no genoma após o códon de início da β 1-amastina, impedindo a tradução. Esta estratégia é mais interessante que a anterior, pois não há o risco de interrupção do gene adjacente, o que poderia influenciar o fenótipo observado após a deleção. Esta estratégia foi possível, e escolhida pelo fato dessas regiões diferirem bastante do restante do genoma, inclusive de outras amastinas.

As regiões para recombinação homóloga aos cromossomos 26 e 32 foram amplificadas de DNA genômico de CL Brener e clonadas no vetor UPPacDN. O gene de resistência a Puromicina foi substituído pelo gene de resistência a Neomicina ou pelo gene de resistência a Higromicina B. Estas etapas de clonagem foram realizadas pela aluna Laiane Lemos. Desta forma, foram obtidos os vetores UP32PacDN32, UP32NeoDN32, UP32HigroDN32, UP26PacDN26, UP26PNeoDN26 e UP26HigroDN26. A figura 19 mostra o esquema do cassete de deleção tanto para a amastina presente no cromossomo 26 como para as presentes no cromossomo 32.

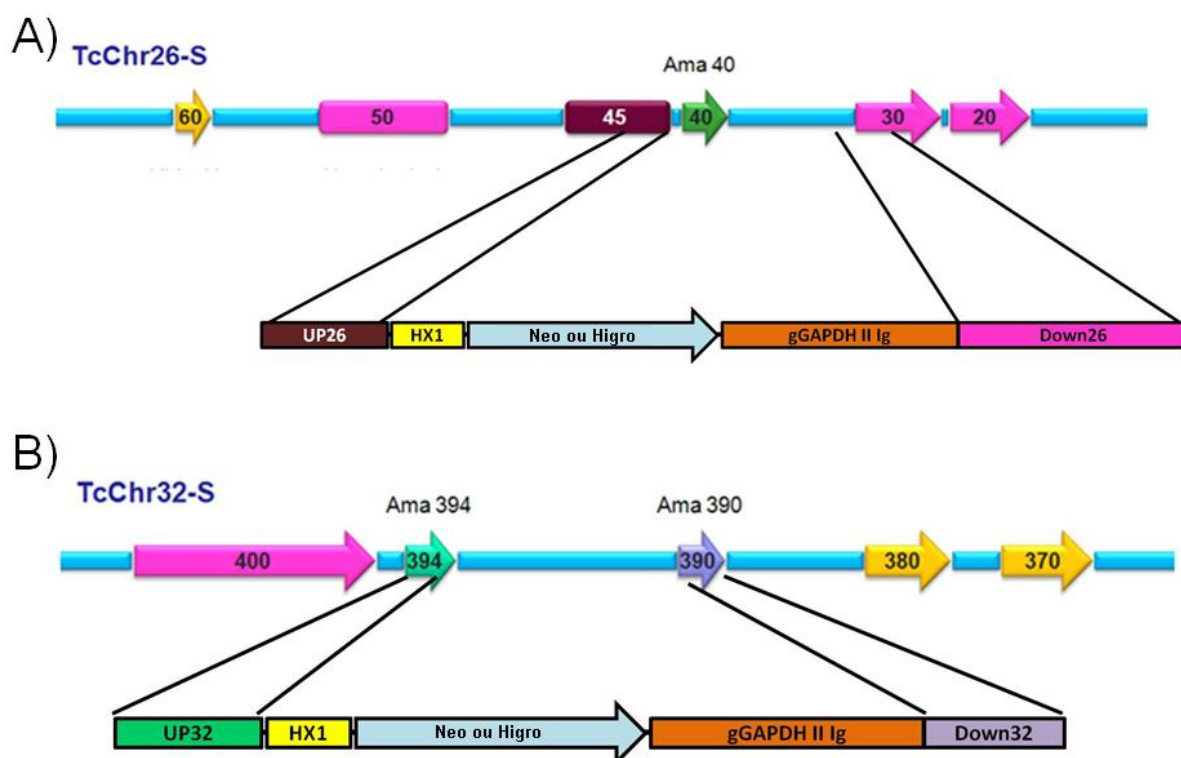


FIGURA 19 – ESQUEMA REPRESENTATIVO DA ESTRATÉGIA DE CONSTRUÇÃO DO VETOR PARA DELEÇÃO DE δ -Ama40 e β -AMASTINAS POR RECOMBINAÇÃO HOMÓLOGA

FONTE: O autor (2014)

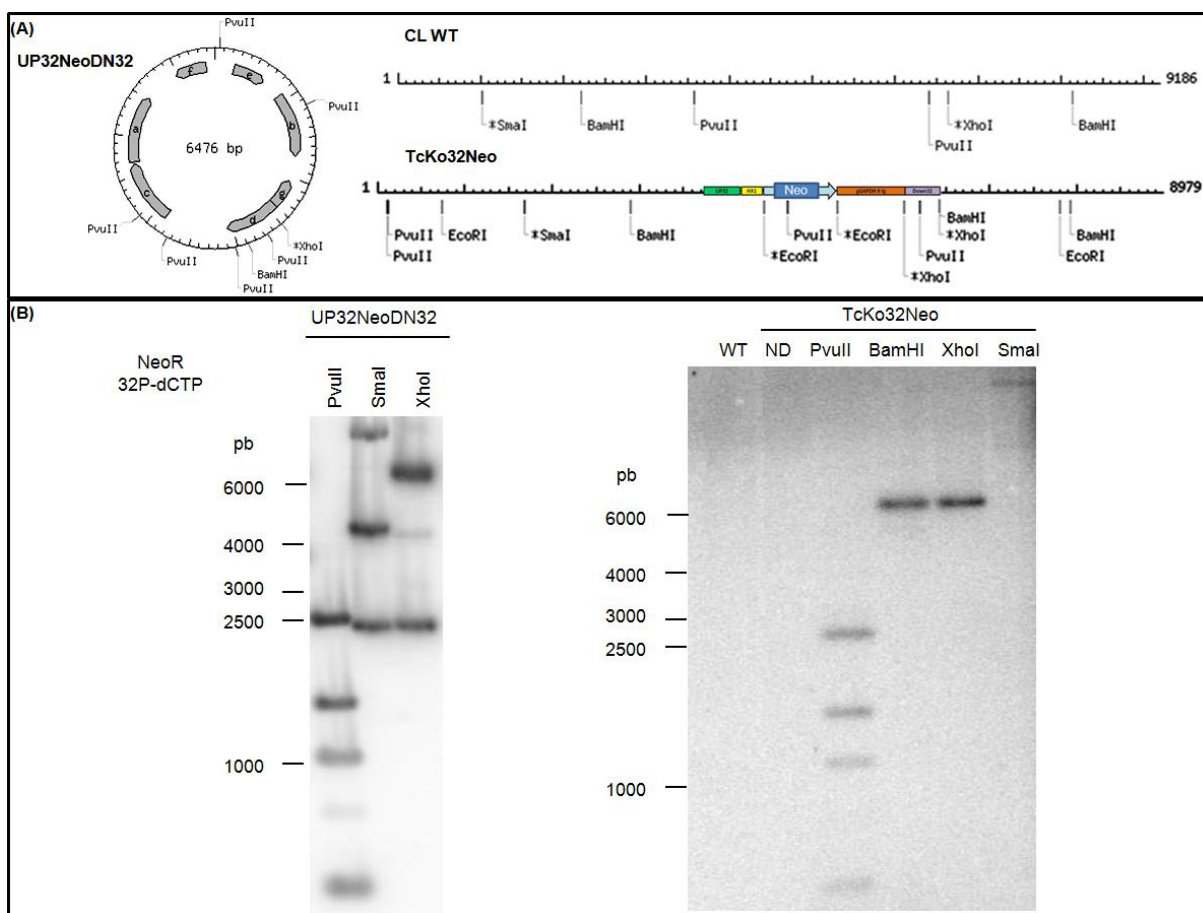
NOTA: O esquema mostra os trechos do genoma de *T. cruzi*, CL Brener que contém as sequências de δ -Ama40 e β -amastinas (setas verdes), presentes nos cromossomos 26 e 32. Para a construção dos vetores as regiões indicadas foram amplificadas e clonadas em vetor contendo resistência a neomicina (Neo) ou higromicina (Histro). Os vetores construídos foram digeridos com *HindIII* e *BamHI* ou amplificados por PCR com os iniciadores apropriados e o produto foi inserido no parasito a fim de se fazer a deleção destas amastinas.

A partir desse momento serão mostrados apenas os resultados referentes a deleção das β -amastinas devido aos avanços obtidos. O cassete de resistência a G418 para deleção por recombinação homóloga foi digerido, submetido a eletroforese, purificado do gel de agarose, eletroporado em formas epimastigotas da cepa CL Brener. A população resistente a G418 foi obtida (TcKo32Neo) e o DNA genômico dessa cultura foi extraído, digerido com as enzimas de restrição *PvuII*, *BamHI*, *XhoI* e *SmaI*, e analisado por *Southern blot*, assim como o DNA do vetor UP32NeoDN32 (figura 20). As membranas foram hibridizadas com sonda marcada com ^{32}P -dCTP para o gene de resistência a Neomicina (Neo^R) e o perfil obtido para

a população resistente foi similar ao obtido para as digestões do vetor utilizado para gerar TcKo32Neo, mostrando que o cassete não foi recombinado (figura 20).

Para a digestão com *Sma*I, o perfil de bandas reconhecidas pela sonda corresponde ao do plasmídeo não digerido, já que não existem sítios de restrição para esta enzima. Ela foi escolhida para tal diagnóstico por existir um sítio no genoma, a jusante da região de recombinação. Para a enzima *Xho*I há somente um sítio de restrição no plasmídeo, devendo linearizar o plasmídeo em 6476 pb. A banda inferior se deve provavelmente à digestão parcial, e corresponde ao plasmídeo supercircular, com tamanho similar ao encontrado para o DNA não digerido (canaleta de digestão com *Sma*I). Já a enzima *Pvu*II cliva o DNA na sequência do gene de resistência a neomicina, utilizado como sonda. O perfil de bandas obtido para a hibridização corresponde ao perfil de restrição do plasmídeo, como mostrado na figura 20c. Estes dados mostram que o cassete para deleção não foi integrado ao genoma e que o plasmídeo estava sendo mantido pelo parasito de forma episomal. Estes resultados indicam que os parasitos foram obtidos a partir de DNA circular contaminantes na transfecção.

Foram realizadas também, a partir de DNA genômico das mesmas populações resistentes (TcKo32Neo), hibridizações com sonda para a ORF de β 1-amastina, porém o perfil de bandas hibridizadas mostrou-se bastante similar ao obtido para a hibridização utilizando DNA do parasito selvagem, indicando que não houve recombinação do cassete neste locus (dados não mostrados). Perfil semelhante, de manutenção do plasmídeo episomal e não recombinação do cassete na região correta, foi encontrado para população resistente obtida da transfecção com UP32HigroDN32 e UP26NeoDN26 (dados não mostrados).



(C)

	UP32-Neo-DN32 Vetor episomal (perfil de restrição)	UP32-Neo-DN32 Vetor episomal (perfil de hibridização)	UP32-Neo-DN32 Cassete integrado após hibridização
BamHI	6476 pb	6476 pb	3520 pb
PvuII	2433, 1507, 1109, 717, 360, 350 pb	1507 e 1109 pb	4564 e 1507 pb
XhoI	6476 pb	6476 pb	Sítios no genoma

FIGURA 20 – ANÁLISE DA INTEGRAÇÃO DOS CASSETES NO GENOMA DE *T. cruzi*

FONTE: O autor (2014)

NOTA: Esquema de integração nos locus de β -amastinas do cromossomo 32. (A) Mapa de restrição esperado para integração do cassete para nocaute e tamanho dos fragmentos esperados na hibridação do *Southern blot* com *Neo^R*. (B) Perfil de hibridação dos transfectantes obtidos. Após a transfecção dos vetores para deleção das amastinas e seleção, foi extraído DNA genômico das populações resistentes a G418. Como controle foi usado a população de CL Brener selvagem. Estas amostras de DNA genômico foram digeridas com as enzimas indicadas na figura, corridas em gel de agarose e transferidas para membrana HybondN+. As amostras WT e ND correspondem ao DNA genômico não digerido do parasito selvagem ou da população indicada respectivamente. Estas foram hibridizadas com sonda marcada com ³²P-dCTP para o gene de resistência a Neomicina. (C) Tamanhos esperados em pb para cada digestão simples.

A fim de reduzir as chances de contaminação com o vetor íntegro na transfecção, partimos para uma nova estratégia de obtenção do DNA para a transfecção. Nessa estratégia, o cassete UP32NeoDN32 foi amplificado por PCR, tratado com *DpnI* para remoção do DNA molde, e transfectado em formas epimastigotas do clone CL Brener de *T. cruzi*. A população resistente a G418 foi clonada por diluição seriada, tendo sido obtidos oito clones para TcKo32Neo em CL Brener. A figura 21 mostra os resultados obtidos da análise desses clones por *southern blot*. Como se pode observar, o perfil de hibridização para todos os clones obtidos é compatível com inserção do cassete UP32NeoDN32 no genoma do parasito no local esperado. Isto pode ser inferido já que um dos sítios de *PvuII* está dentro do gene de resistência a Neomicina, enquanto os outros dois estão no genoma, um a montante e outro a jusante da construção inserida, produzindo fragmentos de 4564 e 1507 pb, com tamanhos distintos do vetor não integrado (1507 e 1109 pb).

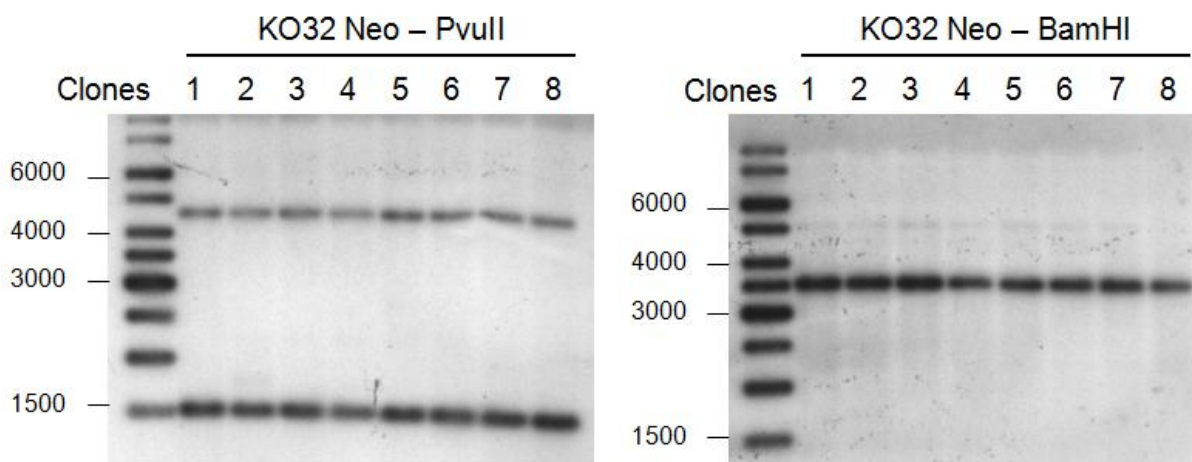


FIGURA 21 – ANÁLISE DA INTEGRAÇÃO DO CASSETE UP32-Neo-DN32 NO LOCUS DAS β -AMASTINAS PRESENTES NO CROMOSSOMO 32

FONTE: O autor (2014)

NOTA: Após a transfecção dos cassetes para deleção das amastinas, seleção das populações resistentes e clonagem por diluição seriada, os clones obtidos foram expandidos e foi realizada a extração do DNA genômico. Estas amostras foram digeridas com as enzimas indicadas na figura, corridas em gel de agarose e transferidas para membrana HybondN+. Estas foram hibridizadas com sonda marcada com ^{32}P -dCTP para o gene de resistência a Neomicina. O perfil de bandas esperado para os vetores e para a integração destes ao genoma está indicado na figura 20c.

A banda detectada na digestão com *Bam*HI tem o tamanho esperado para integração do cassete considerando a retenção do sítio de *Bam*HI que estava presente no iniciador (32DNRevBaXb). Uma nova hibridização com sondas adicionais, além de reações de PCR seguidas de sequenciamento, esclarecerá a integração do cassete no locus correto no genoma. Apesar da necessidade de confirmação, este resultado é promissor, e mostra vantagem da técnica utilizada na segunda estratégia de nocaute. Quando foi utilizada amplificação do cassete seguida de tratamento com *Dpn*I, o perfil de hibridização (figura 21) foi diferente do obtido quando o cassete foi purificado por digestão seguida de eletroforese e excisão da banda, que mostrou um perfil de manutenção do plasmídeo de forma episomal (figura 20). Paralelamente a esses esforços, estão sendo geradas linhagens simples nocaute para as amastinas presentes no cromossomo 26.

Uma vez confirmada a deleção de um dos alelos por *Southern blot* e PCR, o próximo passo será a análise da expressão das amastinas deletadas. De acordo com o resultado obtido prosseguiremos para a remoção do segundo alelo utilizando o cassete UP32HigroDN32 ou UP26HigroDN26 (já construídos) e caracterização fenotípica das linhagens a serem obtidas.

Em resumo, parte da caracterização funcional das diferentes subfamílias de amastinas de *T. cruzi* foi realizada e a avaliação fenotípica da perturbação da expressão das amastinas pela superexpressão e/ou nocautes está em andamento. Primeiramente, foram analisadas as sequências de amastinas e pôde-se observar, apesar de alta variabilidade, uma topologia comum e alta conservação dos aminoácidos que constituem a assinatura das amastinas. Todas as sequências de amastinas dos genomas de CL Brener, Sylvio X-10 e Esmeraldo foram avaliadas quanto à filogenia, quando foram divididas em δ -amastinas, δ -Ama40, β 1- e β 2-amastinas. Quanto à organização genômica, para as δ -amastinas foi encontrado um maior número de cópias em nossa análise, e que estão presentes em maior número de bandas cromossômicas do que mostra a montagem genoma de CL Brener. Para as demais subfamílias, os dados encontrados parecem confirmar o previsto no banco de dados. A avaliação da abundância de mRNA por *northern blot* mostra que

as δ -amastinas estão mais expressas em amastigotas, enquanto as β -amastinas, em epimastigotas. Finalizando a caracterização molecular, as diferentes subfamílias estão localizadas em regiões diferentes, mas sempre associadas a membranas.

Para a caracterização fenotípica foram obtidos clones expressando as diferentes subclasses de amastinas em fusão com GFP em cepa G e clone Sylvio X-10. Também foram obtidos clones com um alelo deletado para ambas as β -amastinas. São necessários mais experimentos a fim de concluir o efeito destas perturbações na expressão de amastinas em *T. cruzi*. Não foi possível desenhar uma estratégia de nocaute para δ -amastinas pois a única maneira disponível é o nocaute clássico, por recombinação homóloga, o que é bastante complicado no caso de genes multicópias e com regiões repetidas. Devido a estes obstáculos, a segunda parte deste projeto tem como finalidade a padronização de uma nova ferramenta de geração de nocautes em *T. cruzi*, o sistema DiCre.

6.2 PARTE 2: DESENVOLVIMENTO DO SISTEMA DICRE PARA OBTENÇÃO DE NOCAUTES CONDICIONAIS EM *TRYPANOSOMA CRUZI*

A fim de ampliar a limitada “caixa de ferramentas de biologia molecular” disponível para manipulação gênica de *T. cruzi* escolhemos investir na geração de um sistema de nocaute condicional, o sistema DiCre. Como mencionado anteriormente, nesse sistema a regulação é pós-traducional pela ação do ligante rapamicina, logo não há necessidade de utilização de cepas previamente modificadas para expressão regulada, como foi utilizado anteriormente para diminuir o efeito tóxico da recombinase CRE por meio do uso do sistema T7RNAPolimerase/ Repressor Tetraciclina (BARRETT *et al.*, 2004; SCAHILL *et al.*, 2008). Nesse contexto há uma redução na necessidade de marcas de seleção.

Para possibilitar o desenvolvimento deste sistema em *T. cruzi*, foram obtidos os vetores pODiCre Higro (figura 8) e pROCK FEKO_Pur Neo (figura 7), o primeiro desenhado para expressão das duas porções da recombinase DiCre, e o segundo o

vetor teste, o qual irá sofrer recombinação sítio dirigida em sítios *loxP* por recombinase DiCre. Estes vetores foram usados para obtenção de populações duplo-resistentes, as quais foram clonadas por diluição limitante. Após quatro semanas de cultivo, foram obtidos três clones contendo o plasmídeo pODiCre Higo, três clones contendo pROCK FEKO_Pur Neo e dois clones contendo ambos os plasmídios. Estes clones foram identificados como: TcDiCre1, TcDiCre2 e TcDiCre3, TcFEKO1, TcFEKO2, TcFEKO3, TcDF1 e TcDF2, respectivamente.

Uma vez geradas linhagens de parasitos carregando pROCK FEKO_Pur Neo (resistente a G418), foi testada a expressão do segundo gene de resistência (Pur^R) presente na construção, e que está flanqueado por sítios *loxP*. Como se pode observar na figura 22, tanto na presença quanto na ausência do antibiótico, a população transfectada com o plasmídeo linearizado, chamada aqui de FP, manteve seu crescimento *in vitro* praticamente inalterado na presença da droga. Já CL Brener selvagem, na presença do antibiótico houve uma drástica inibição do crescimento. Este resultado sugere que o *T. cruzi* é capaz de expressar o gene de resistência a puromicina (PAC – ou PurR) a partir do vetor pROCK FEKO_Pur Neo. Além disso, este resultado mostra que PAC talvez possa ser utilizado como marca de seleção em experimentos futuros.

De acordo com a estratégia de clonagem do cassete de PAC, a enzima de resistência a puromicina é expressa em fusão a enzima HSVTK (Timidina Kinase de *Herpes Simplex Virus*), que foi utilizada de forma eficiente como gene suicida em *T. brucei* (SCAHILL *et al.*, 2008). Contudo, quando testamos a sensibilidade de formas epimastigotas a ganciclovir nessas populações, as formas epimastigotas foram insensíveis ao tratamento (dados não mostrados). Estes dados estão de acordo com observações anteriores que mostraram que por razões desconhecidas epimastigotas de *T. cruzi* são insensíveis a pró-droga (BUCKNER *et al.*, 1997).

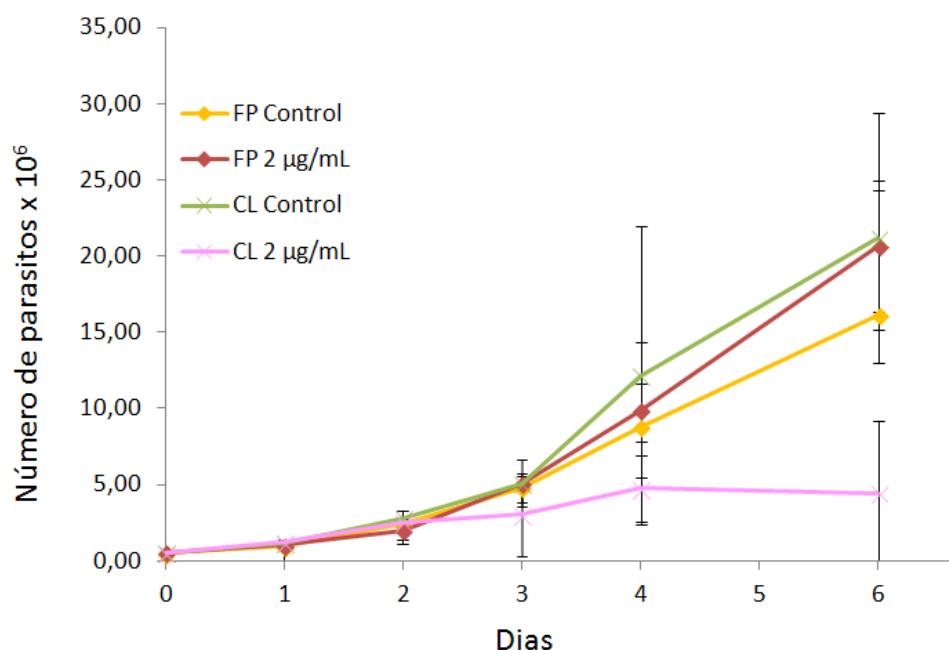


FIGURA 22 – SENSIBILIDADE DE FORMAS EPIMASTIGOTAS DE *T. cruzi* A PUROMICINA

FONTE: O autor (2014)

NOTA: *T. cruzi*, da cepa CL Brener foram transfectados com pROCK FEKO_Pur Neo e selecionados com G418. A população resistente (FP) e o clone selvagem CL Brener (CL) foram inoculadas na presença e ausência de 2 µg/mL de puromicina. A densidade de células foi determinada em microscópio, por contagem em câmara de Neubauer. As culturas foram diluídas cada vez que atingia a densidade de 2.10^7 parasitos/mL. Os resultados estão expressos como média \pm SD (n = 3).

Para um diagnóstico rápido da presença da construção pODiCre Higro nos transfectantes, utilizou-se PCR com o par de iniciadores Cre60For e Cre60Rev usando como molde o DNA genômico dos clones. Com este ensaio obteve-se o tamanho do amplicon correspondente ao esperado (dados não mostrados). Uma vez confirmada a presença dos plasmídios no genoma, a integração das construções e localização no genoma foi avaliada por PFGE, seguido de *Southern blot* e hibridação com sondas específicas (figura 23).

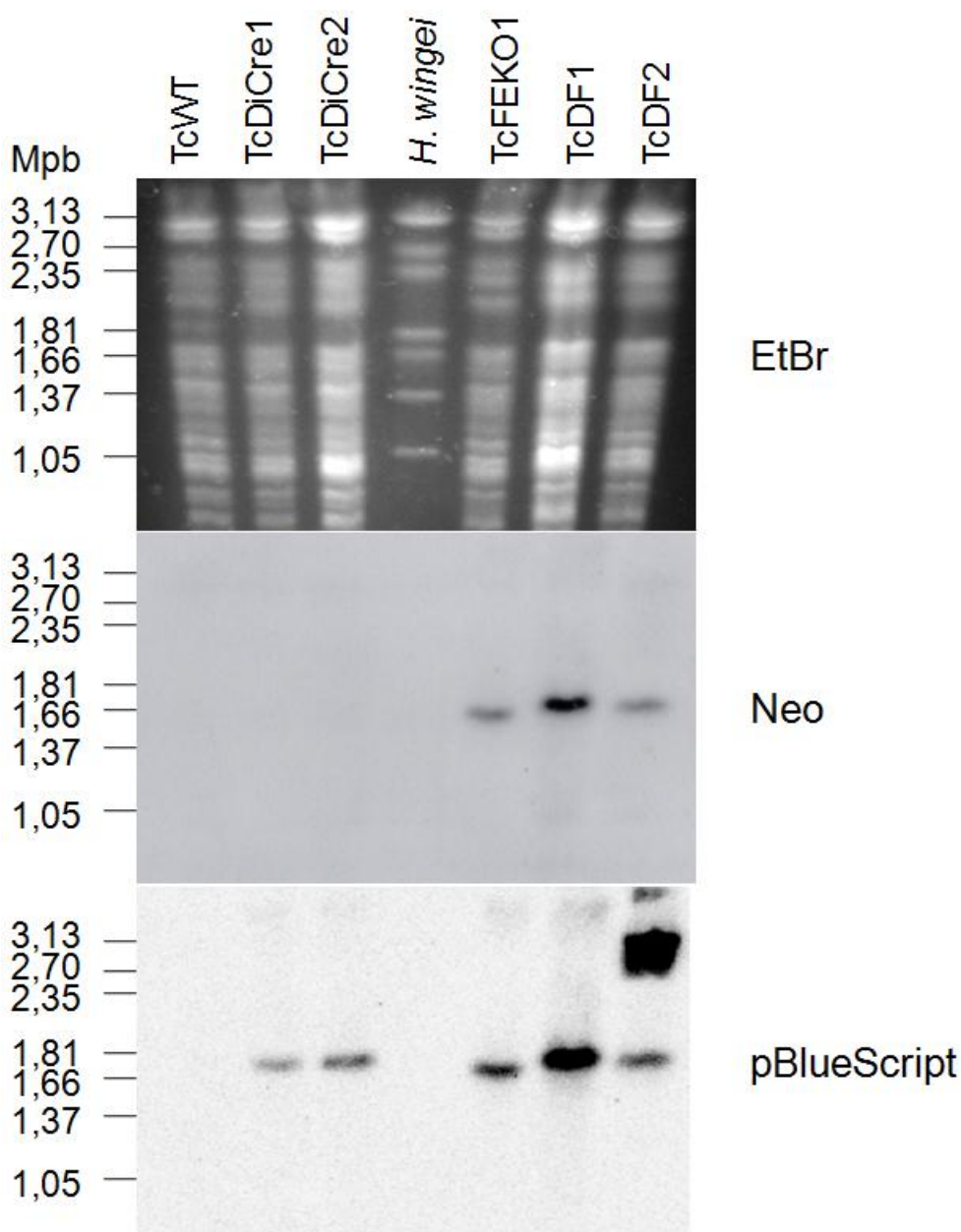


FIGURA 23 – PERFIL ELETROFORÉTICO POR PFGE E DE HIBRIDIZAÇÃO DE DNA CROMOSSOMAL COM SONDA RADIOATIVA

FONTE: O autor (2011)

NOTA: O primeiro painel mostra o perfil obtido por eletroforese de campo pulsátil, corado com brometo de etídeo, de DNA cromossômico obtido das amostras TcWT, TcDiCre1, TcDiCre2, *H. wingei* (usado para marcação de peso molecular) TcFEKO1, TcDF1 e TcDF2. Ao meio está mostrado a integração do vetor pROCK FEKO_Pur Neo ao genoma do parasito através do perfil de hibridização do DNA cromossômico com sonda obtida da amplificação por PCR do gene de resistência a Neomicina com iniciadores NeoF e NeoR. Mais abaixo está a mesma membrana hibridizada com sonda para pBlueScript, obtida por digestão de pROCK GFP Higro com *PvuII*, seguida de purificação após separação em gel de agarose.

Como é possível observar na figura 23, houve hibridização das sondas de “NeoR” e “pBluescript” em um tamanho de aproximadamente 1,66 Mpb. De acordo com o cariótipo do clone CL Brener de *T. cruzi*, por volta deste tamanho, há o cromossomo de número 15, com 1,64 Mpb, que já foi marcado anteriormente com sonda para o gene da α -tubulina (VARGAS, PEDROSO E ZINGALES, 2004). Foram observadas bandas de tamanhos próximos a este para integração do vetor pROCK em *T. cruzi*, cepa Tulahuén (DAROCHA *et al.*, 2004b; PIRES *et al.*, 2008). Estes dados mostram que houve integração do vetor ao genoma do parasito.

6.2.1 Teste de recombinação por DiCre recombinase

Como descrito acima, os clones duplo resistentes (TcDF) contêm os plasmídios pROCK FEKO_Pur Neo e pODiCre Higro, sendo assim eles possuem os cassetes para a expressão das porções de CRE recombinase, Cre59 e Cre60 fusionados aos domínios FKBP12 e FRB, respectivamente, que se mantêm aproximados na presença de rapamicina, formando a Cre recombinase dimerizável, ou seja, a topoisomerase ativa. Estes parasitos também possuem um fragmento de DNA flanqueado por sítios *loxP*, ou seja, um segmento de DNA *floxed* que será utilizado no diagnóstico da recombinação.

Na figura 24A está ilustrado o mapa de restrição do vetor pROCK FEKO_PurNeo integrado no genoma, antes e após a recombinação por DiCre. Anteriormente a recombinação, pode-se observar que com a digestão com *PvuII*, que reconhece um sítio no gene de resistência a neomicina, e outros dois sítios, um a 4143 pb a montante do sítio em Neo^R e outro a 1503 pb a jusante do sítio em Neo^R, são detectadas duas bandas de tamanhos 4143 e 1503 pb. Se houver recombinação pela atividade da DiCre recombinase, ocorrerá o aparecimento de uma banda de 2162 pb na mesma proporção em que reduzirá a intensidade da banda de 4143 pb, sendo que a banda de 1503 permanecerá inalterada. Nesta estratégia a deleção por recombinação é analisada por RFLP (Restriction Fragment Length Polymorphism).

O clone transfectado TcDF1, e o clone selvagem de CL Brener (CL WT) foram expostos a 0, 10 e 50 nM de rapamicina por 12 horas (figura 25). Como esperado não houve hibridização da sonda com o DNA do parasito selvagem. Para o clone TcDF1 não induzido pode-se observar que houve hibridização somente nos tamanhos previstos para o vetor não recombinado. Este resultado mostra que não houve atividade residual da DiCre recombinase na ausência do indutor, ou que está reduzida a níveis não detectáveis pela técnica utilizada. Com a indução é possível observar que, para estas amostras, houve hibridização para os tamanhos previstos para o vetor não recombinado e ocorreu também o aparecimento da banda no tamanho previsto para o vetor recombinado, evidenciando o funcionamento do sistema DiCre/*loxP* em *T. cruzi* nas condições testadas, mesmo que com baixa eficiência.

Para que a recombinase atue na excisão do trecho de DNA flanqueado por *loxP*, a proteína precisa ser direcionada para o núcleo. Nas sequências de Cre59 e Cre60 são encontrados sinais de localização nuclear (NLS) do vírus SV40 (PKKKRKV) (JULLIEN *et al.*, 2003). Este mesmo sinal já foi utilizado anteriormente para direcionar para o núcleo a Cre59 e Cre60 em *Toxoplasma* sp. (ANDENMATTEN *et al.*, 2013) e T7 RNA polimerase em *T. brucei* (WIRTZ *et al.*, 1999). Há evidências que este sinal de endereçamento também funciona em *T. cruzi*, pois a mesma T7 RNA polimerase contendo NLS foi utilizada para construção de sistema de expressão regulada (DAROCHA *et al.*, 2004a; TAYLOR E KELLY, 2006).

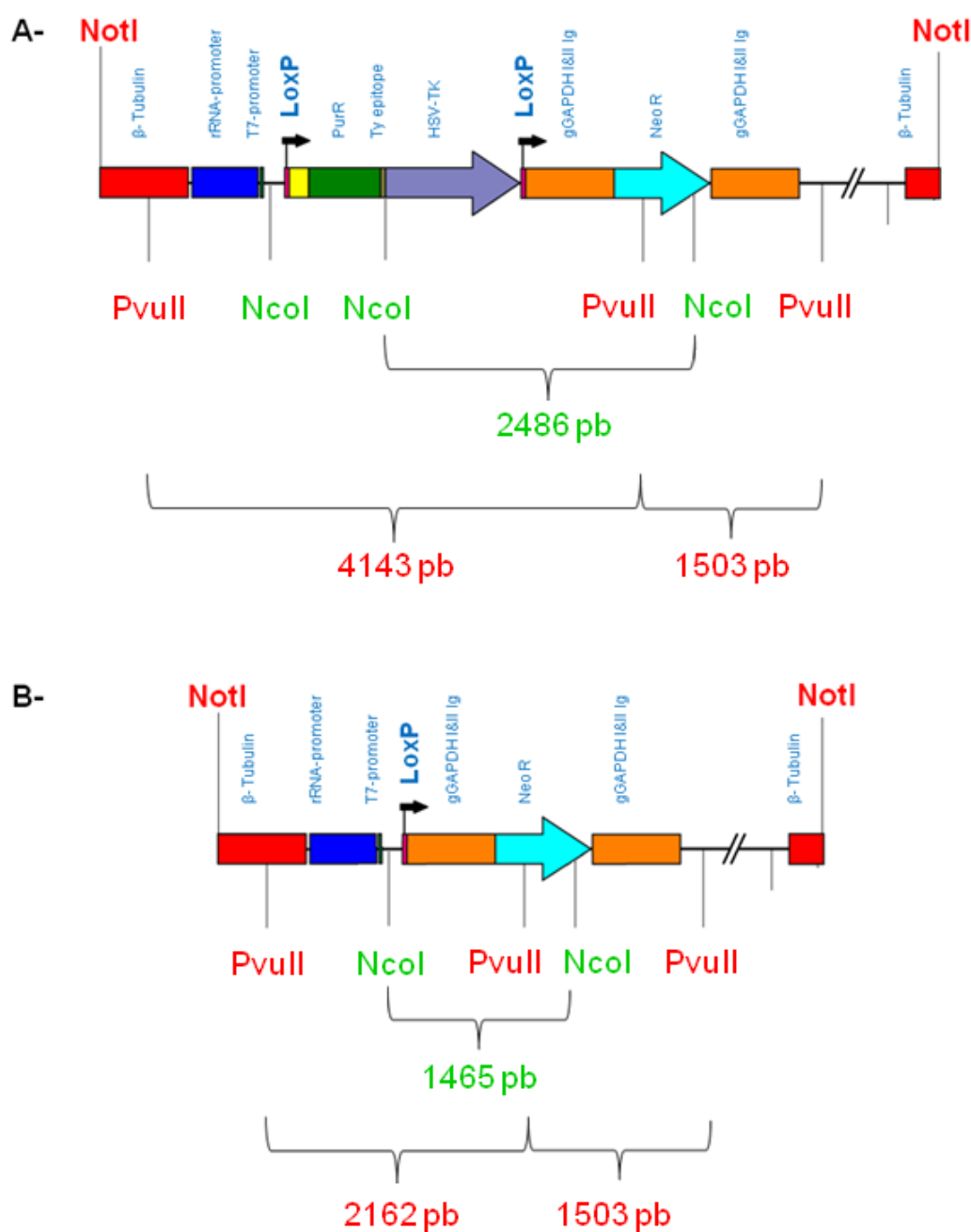


FIGURA 24 – ESQUEMA REPRESENTATIVO DO VETOR pROCK FEKO_Pur Neo ANTES E APÓS RECOMBINAÇÃO MEDIADA POR CRE RECOMBINASE

FONTE: O autor (2014)

NOTA: O esquema mostra o vetor pROCK FEKO_Pur Neo linearizado com NotI. O plasmídeo contém PAC (Puromycin N-acetyl-transferase) fusionada com HSV-TK (Timidina quinase de Herpes Simplex vírus – gene suicida) flanqueado por sítios *loxP* (→). Após a recombinação mediada por CRE recombinase, há a deleção do cassete construção entre os sítios *loxP* e de um dos sítios, restando somente um deles. Estão indicados na figura os sítios de restrição para *PvuII* e o número de pares de base entre estes sítios.

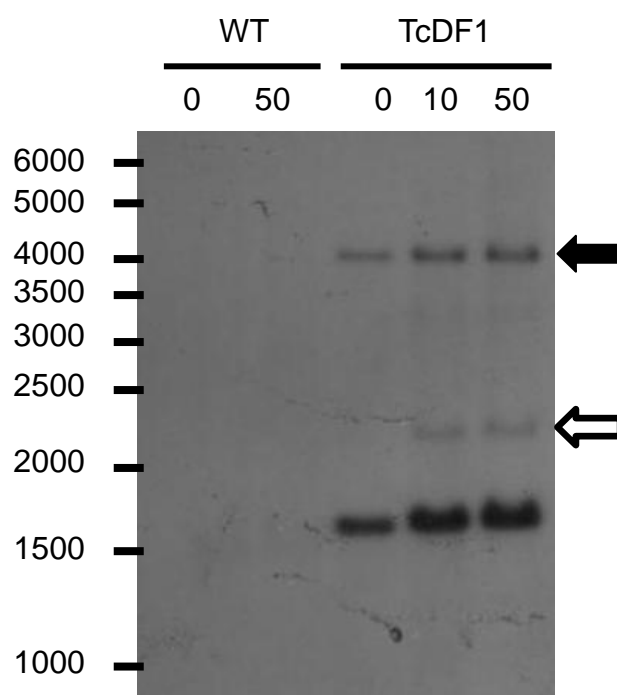


FIGURA 25 – SOUTHERN BLOT DO TESTE DE RECOMBINAÇÃO MEDIADO POR CRE RECOMBINASE INDUZIDA POR RAPAMICINA

FONTE: O autor (2014)

NOTA: A figura mostra o resultado obtido com 12 horas de indução do clone TcDF1 a 0, 10 e 50 nM de rapamicina. Como controle foi utilizado o parasito selvagem (WT), *T. cruzi*, cepa CL Brener na ausência e presença de 50 nM de rapamicina, também por 12 horas. A seta cheia indica a provável banda de 4143 pb, correspondente ao tamanho previsto do vetor não recombinação por CRE recombinase. A seta vazia indica a provável banda de 2162 pb, que corresponde ao tamanho previsto para recombinação mediada por CRE. A banda mais a baixo é provavelmente a banda de 1503 pb, que não sofre alteração após atividade da enzima.

A fim de elevar a taxa de recombinação mediada por recombinase CRE, foram realizados testes com variação dos tempos de indução e concentração de rapamicina, mostrados na figura 26. Estes resultados mostram que a variação de tempo e dosagem de rapamicina não aumentou a eficiência de recombinação da CRE dimerizável para níveis desejáveis para obtenção de nocautes, já que a população obtida é heterogênea, ou seja, composta por parasitos que sofreram ou não sofreram recombinação.

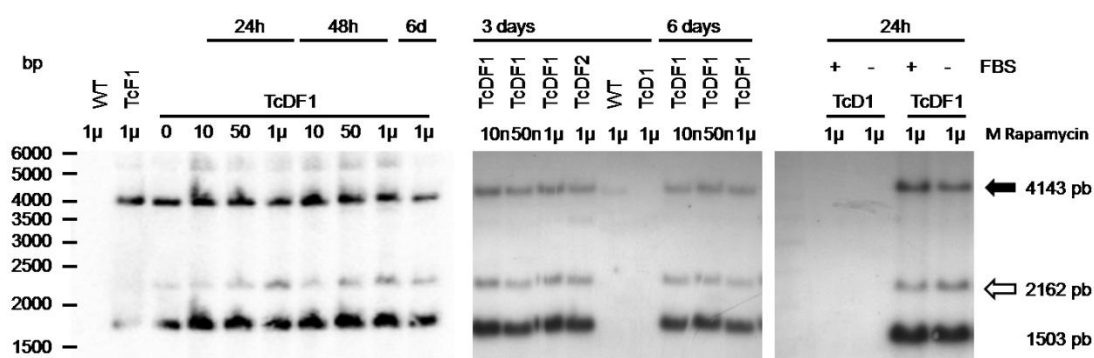


FIGURA 26 – ANÁLISE POR SOUTHERN BLOT DO TESTE DE RECOMBINAÇÃO MEDIADO POR CRE RECOMBINASE INDUZIDA POR RAPAMICINA EM DIFERENTES CONCENTRAÇÃO E TEMPOS DE INDUÇÃO

FONTE: O autor (2014)

NOTA: (A) O painel mostra o vetor pROCK FEKO_Pur Neo linearizado antes e após a recombinação. Após a recombinação mediada por CRE Recombinase, um fragmento de aproximadamente 2000 pb flanqueado por sítios *loxP* (→) é deletado. O evento pode ser facilmente detectado por polimorfismo no comprimento dos fragmentos de restrição (RFLP) usando restrição com *PvuII* seguida de Southern Blot com sonda marcada com ^{32}P –dCTP. (B) Resultado obtido da indução dos clones TcF1, TcD1, TcDF1 e TcDF2 com diferentes concentrações de rapamicina por diferentes períodos. Como controle foi utilizado o parasito selvagem (WT), *T. cruzi*, cepa CL Brener. (C) A seta cheia indica a provável banda de 4143 pb, correspondente ao tamanho previsto do vetor não recombinado por recombinase CRE. A seta vazia indica a provável banda de 2162 pb, que corresponde ao tamanho previsto para recombinação mediada por CRE. A banda mais a baixo é provavelmente a banda de 1503 pb, que não sofre alteração após atividade da enzima.

A eficiência de recombinação da DiCre varia em diferentes organismos. Quando o sistema DiCre foi desenvolvido em fibroblastos Rat2, com 10 nM de um análogo a rapamicina foi obtida 75% de recombinação após 48h e ~100% após 72 horas (JULLIEN *et al.*, 2003). Quando transferido para camundongos, a maior taxa de recombinação obtida, medida por *Southern blot* usando DNA de células hepáticas, foi de 40%, após tratamento dos animais por 10 dias cinco injeções diárias de 40 mg/Kg de rapamicina (JULLIEN *et al.*, 2007).

Em *T. brucei* a recombinase CRE foi utilizada pela primeira vez na forma íntegra para retirar uma sequência terminadora da transcrição ($\text{TERM}_{\text{GPEET}}$) flanqueada por *loxP*, que após ser retirada permitia a expressão de GFP. Apesar de ter sua expressão regulada por promotor induzível por tetraciclina (pLEW100Cre), a recombinase CRE apresentou atividade passados alguns dias de cultivo mesmo sem a adição do indutor. Foi avaliado por citometria de fluxo que a fluorescência das

culturas aumentava com o passar dos dias, apresentando de 22 a 50% de fluorescência durante as cinco semanas de cultivo sem indutor. Este dado foi confirmado por PCR, obtendo um amplicon do tamanho esperado para a recombinação dos sítios *loxP* (BARRETT *et al.*, 2004). Ainda em *T. brucei*, a fim de melhorar a regulação, e consequentemente a atividade de recombinação, outro trabalho modificou o vetor pLEW100Cre vetor e usou em formas sanguíneas, na qual a expressão é diminuída em 10X.

Outra alternativa usada neste trabalho foi a transfecção transiente do plasmídeo não modificado em uma população já contendo os sítios *loxP* flanqueando um gene de resistência a higromicina fusionada ao gene HSVTK. Após a transfecção, a população foi clonada na presença de ganciclovir (GCV) para seleção dos clones que tornaram resistentes a essa droga após a atividade de recombinação. A taxa de clones resistentes a GCV foi de 3%. Apesar de baixa, todos os clones obtidos são sensíveis a higromicina, e não apresentaram diminuição na taxa de crescimento. Esta alternativa ainda é vantajosa por não ser necessário o uso de linhagens com o sistema de indução a tetraciclina, que já são resistentes a dois antibióticos, e podem ser usadas para retirada de marcas de seleção. Contudo, a etapa de clonagem é indispensável para obtenção de populações homogêneas (SCAHILL *et al.*, 2008).

Desta forma, como neste trabalho foram obtidas populações mistas de recombinação, a fim de isolar parasitos 100% recombinados, os parasitos foram clonados após a indução com rapamicina. No diagnóstico da recombinação, o DNA genômico dos clones foi digerido com uma segunda enzima para confirmar o perfil de recombinação. Na figura 24, o esquema mostra os tamanhos esperados para a hibridização das membranas com sonda para o gene de resistência a Neomicina após digestão com *NcoI*.

Para este experimento as populações TcDicre, TcFeko, TcDF, além do parasito não transfectado (CL WT), foram induzidas, ou não, por 48h com 1 mM de rapamicina e clonadas na ausência ou presença do indutor na mesma concentração. Foram obtidos seis clones da indução de TcFeko que após serem digeridos com

NcoI não apresentaram perfil de recombinação *loxP* dirigida, o que era esperado devido a ausência da expressão da DiCre. Também como era esperado, os 12 clones obtidos da população TcDF não induzida apresentaram o perfil de não recombinado para as duas enzimas usadas (dados não mostrados). A figura 27 mostra os resultados obtidos para os clones de TcDF pré-induzida dor 48 horas com 1 mM de Rapamicina e clonadas na presença ou ausência deste indutor.

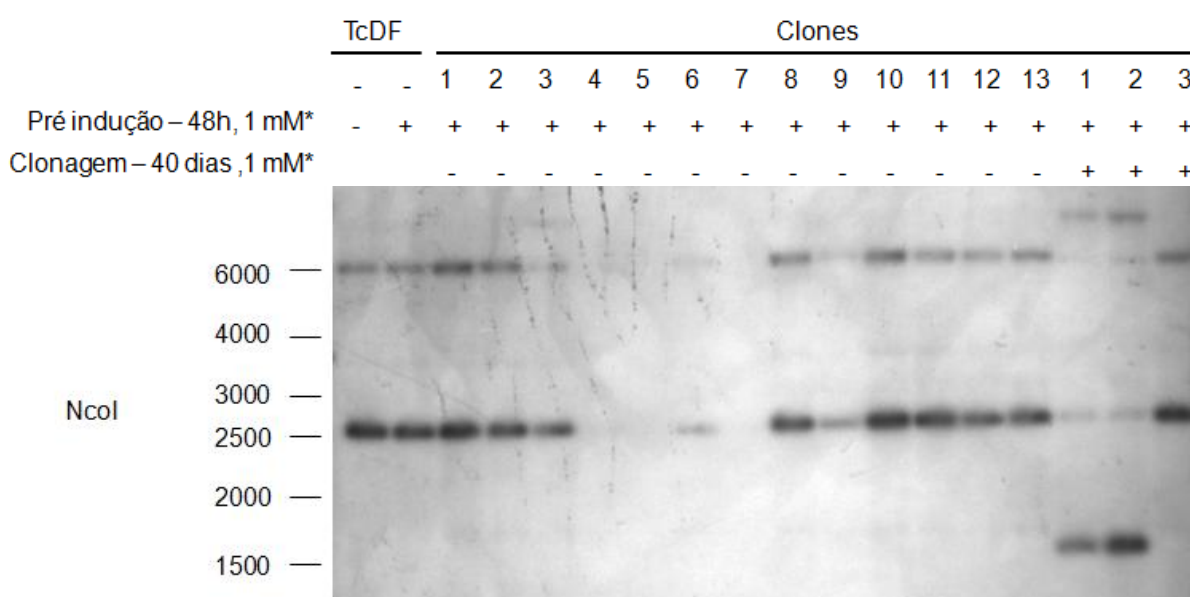


FIGURA 27 – TESTE DE RECOMBINAÇÃO MEDIADO POR DICRE RECOMBINASE APÓS CLONAGEM.

FONTE: O autor (2014)

NOTA: Resultado obtido da pré-indução da população TcDF com 1 mM de rapamicina por 48 horas seguida de clonagem na presença ou ausência da mesma concentração do indutor. As culturas foram expandidas e foi extraído DNA genômico, que foi digerido com a enzima de restrição *NcoI*. Após correr as digestões em eletroforese em gel de agarose e transferir para membranas de nylon, estas foram hibridizadas com sonda para o gene de resistência a Neomicina (Neo^R) marcada com $\alpha^{32}\text{P}$ – dCTP. Os tamanhos de bandas previstos são de 2486 pb para não recombinação e 1465 pb para recombinação por recombinase CRE.

Como se pode observar, tanto a população não induzida, como a população induzida, os 13 clones mantidos sem rapamicina apresentaram perfil de não recombinados. Dos três clones obtidos na clonagem mantida na presença de 1 mM de rapamicina, dois deles parecem apresentar perfil de recombinação na digestão com *NcoI*, uma vez que foi detectado o fragmento de interesse. Contudo, este dado

precisa ser repetido uma vez que também detectamos alteração no tamanho da banda de tamanho superior. A fim de delimitar a atividade da DiCre na presença de rapamicina, é necessário avaliar a exposição de baixas concentrações de células, e em tempos prolongados.

Paralelamente, a expressão da DiCre foi avaliada por *western blot* utilizando anticorpo anti-CRE (Santa Cruz). Nas populações testadas, contendo o plasmídeo pODiCre, (TcDF1, TcD1) não foi observada nenhuma banda específica em relação ao controle CL Brener WT (dados não mostrados). Contudo, este resultado não é conclusivo, pois não tínhamos um controle positivo, além disso outros fatores podem ter influenciado, tais como: o fato do anticorpo policlonal utilizado reconhecer uma porção interna da recombinase CRE que pode estar alterada quando a enzima é dividida. A enzima pode também estar expressa em níveis abaixo da detecção do método. Este efeito que pode estar acentuado na ausência de rapamicina, pois quando não associado ao ligante o domínio FKBP funciona como domínio desestabilizador. A fusão deste com proteínas recombinantes já foi utilizada como ferramenta para perturbar a expressão de proteínas de interesse em diversos modelos, como fibroblastos, *Leishmania* sp. e *T. cruzi* (BANASZYNSKI *et al.*, 2006; MA *et al.*, 2012; MADEIRA DA SILVA *et al.*, 2009).

Com os resultados obtidos, pode-se concluir que, mesmo com redução no número de clones obtidos, a manutenção da cultura em rapamicina durante a clonagem aumenta a eficiência de recombinação dos parasitos TcDF que chegou a 66%. Por esta metodologia, a etapa de clonagem é indispensável para a obtenção de populações homogêneas de recombinação, porém os dados preliminares obtidos neste trabalho mostram maior eficiência (66%) em relação a obtida anteriormente (3%) (SCAHILL *et al.*, 2008). Sendo assim este sistema poderá ser utilizado para remoção de marcas de resistência, para nocautes de genes não essenciais, ou deleção de longas regiões cromossômicas.

7 CONCLUSÃO

As sequências das amastinas de *T. cruzi* são bastante divergentes e podem ser estratificadas em mais categorias que o proposto anteriormente. As δ -amastinas estão expandidas em CL Brener, e seu número de cópias está subestimado na montagem atual do genoma. Todas as proteínas avaliadas parecem estar associadas a membranas celulares, podendo estar na superfície do parasito, ou associadas a estruturas internas presentes na fração insolúvel da célula, tratando-se, provavelmente, de membranas internas.

A expressão de δ -amastinas é maior em amastigotas, como foi descrito inicialmente, exceto para cepa G, cujas formas amastigotas extracelulares têm uma maior capacidade invasiva *in vitro*. Diferentemente, a expressão de β -amastinas é mais alta em epimastigotas. Essa alternância de expressão das δ - e β -amastinas aliadas aos seus padrões de localização subcelular distintos sugere uma adaptação da função não apenas no estágio intracelular, mas também ao inseto vetor, quando mais expressa em formas epimastigotas.

Em relação às alterações fenotípicas, foram obtidos parasitos expressando amastinas em fusão com GFP de forma heteróloga e hemi-nocautes de β -amastinas, porém foram realizados testes preliminares de crescimento somente com a primeira estratégia. As alterações fenotípicas dos superexpressores e dos nocautes ainda precisam ser caracterizadas.

Ainda não é possível realizar o nocaute de δ -amastinas ou de β - e δ -Ama40 na mesma população devido a restrição na utilização de ferramentas de biologia molecular. Tentando aumentar as possibilidades de manipulação da expressão gênica, investiu-se no desenvolvimento de uma ferramenta que ainda não foi testada em *T. cruzi*, o sistema de recombinação sítio dirigida utilizando Recombinase DiCre, uma recombinase CRE regulada por rapamicina. As populações contendo o vetor que apresentam dois *loxP* e para expressão de DiCre recombinase foram obtidas. Após uma série de testes de indução com rapamicina, encontramos uma condição

na qual a eficiência de recombinação chegou a 66%, porém é preciso realizar uma etapa de clonagem, o que possibilita o uso desta ferramenta para retirada de marcas de seleção ou nocaute de genes não essenciais. Apesar de ser difícil a sua utilização na geração de nocautes condicionais, o funcionamento deste sistema é um grande avanço, pois eleva drasticamente as possibilidades de manipulação gênica de *T. cruzi*. Mais testes são necessários para padronizar as condições de utilização dessa metodologia em *T. cruzi*.

8 PERSPECTIVAS

Como perspectivas para este trabalho temos:

Concluir as análises fenotípicas dos clones expressando amastinas das diferentes subfamílias em fusão com GFP na cepa G e clone Sylvio X-10. Para isso pretendemos repetir a curva de crescimento acompanhando a fluorescência e a diferenciação em tripomastigotas de cultura envelhecida. Além disso, pretendemos avaliar o perfil de infecção destes clones.

Confirmar a deleção das β -amastinas, obter duplo nocautes e caracterizar o fenótipo destes nocautes. O mesmo procedimento será realizado para a deleção da δ -Ama40.

Confirmar a atividade de recombinação sítio específica do sistema DiCre//oxP em *T. cruzi*. Para isso os clones obtidos nos testes de recombinação serão avaliados por PCR e por RFLP utilizando outras sequências de DNA como sondas. Estes clones também serão avaliados quanto a resistência a puromicina. Também serão realizados novos testes de recombinação para delimitar o tempo e a quantidade de células ideais para obter alta eficiência de recombinação.

9 REFERÊNCIAS BIBLIOGRÁFICAS

- ALLEN, S.; RICHARDSON, J. M.; MEHLERT, A. *et al.* Structure of a complex phosphoglycan epitope from gp72 of *Trypanosoma cruzi*. *J Biol Chem*, v. 288, p. 11093-11105, 2013.
- ALVES, M. J.; MORTARA, R. A. A century of research: what have we learned about the interaction of *Trypanosoma cruzi* with host cells? *Mem Inst Oswaldo Cruz*, v. 104 Suppl 1, p. 76-88, 2009.
- ANDENMATTEN, N.; EGARTER, S.; JACKSON, A. J. *et al.* Conditional genome engineering in *Toxoplasma gondii* uncovers alternative invasion mechanisms. *Nat Methods*, v. 10, p. 125-127, 2013.
- ANDREWS, N. W.; ROBBINS, E.; LEY, V. *et al.* Stage-specific surface antigens during the morphogenesis of *Trypanosoma cruzi*: developmentally regulated expression of a glycosyl-phosphatidylinositol anchored glycoprotein of amastigotes. *Mem Inst Oswaldo Cruz*, v. 83 Suppl 1, p. 561-562, 1988a.
- ANDREWS, N. W.; ROBBINS, E. S.; LEY, V. *et al.* Developmentally regulated, phospholipase C-mediated release of the major surface glycoprotein of amastigotes of *Trypanosoma cruzi*. *J Exp Med*, v. 167, p. 300-314, 1988b.
- ARAUJO, P. R.; BURLE-CALDAS, G. A.; SILVA-PEREIRA, R. A. *et al.* Development of a dual reporter system to identify regulatory cis-acting elements in untranslated regions of *Trypanosoma cruzi* mRNAs. *Parasitol Int*, v. p. 2011.
- ASLETT, M.; AURRECOECHEA, C.; BERRIMAN, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*, v. 38, p. D457-462, 2010.
- ATWOOD, J. A., 3RD; WEATHERLY, D. B.; MINNING, T. A. *et al.* The *Trypanosoma cruzi* proteome. *Science*, v. 309, p. 473-476, 2005.
- AVILA, A. R.; DALLAGIOVANNA, B.; YAMADA-OGATTA, S. F. *et al.* Stage-specific gene expression during *Trypanosoma cruzi* metacyclogenesis. *Genet Mol Res*, v. 2, p. 159-168, 2003.
- BANASZYNSKI, L. A.; CHEN, L.-C.; MAYNARD-SMITH, L. A. *et al.* A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell*, v. 126, p. 995-1004, 2006.
- BARRETT, B.; LACOUNT, D. J.; DONELSON, J. E. *Trypanosoma brucei*: a first-generation CRE-loxP site-specific recombination system. *Exp Parasitol*, v. 106, p. 37-44, 2004.
- BARTHOLOMEU, D. C.; CERQUEIRA, G. C.; LEO, A. C. *et al.* Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res*, v. 37, p. 3407-3417, 2009.
- BATISTA, T. M.; MARQUES, J. T. RNAi pathways in parasitic protists and worms. *J Proteomics*, v. p. 2011.
- BELLOFATTO, V.; CROSS, G. A. Expression of a bacterial gene in a trypanosomatid protozoan. *Science*, v. 244, p. 1167-1169, 1989.
- BENABDELLAH, K.; GONZALEZ-REY, E.; GONZALEZ, A. Alternative trans-splicing of the *Trypanosoma cruzi* LYT1 gene transcript results in compartmental and

- functional switch for the encoded protein. *Molecular Microbiology*, v. 65, p. 1559-1567, 2007.
- BERN, C.; MONTGOMERY, S. P. An estimate of the burden of Chagas disease in the United States. *Clin Infect Dis*, v. 49, p. e52-54, 2009.
- BERRIMAN, M.; GHEDIN, E.; HERTZ-FOWLER, C. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science*, v. 309, p. 416-422, 2005.
- BOCCHI, E. A.; ARIAS, A.; VERDEJO, H. *et al.* The reality of heart failure in Latin America. *J Am Coll Cardiol*, v. 62, p. 949-958, 2013.
- BOLHASSANI, A.; GHOLAMI, E.; ZAHEDIFARD, F. *et al.* Leishmania major: Protective capacity of DNA vaccine using amastin fused to HSV-1 VP22 and EGFP in BALB/c mice model. *Exp Parasitol*, v. 128, p. 9-17, 2011.
- BRINGAUD, F.; GARCIA-PEREZ, J. L.; HERAS, S. R. *et al.* Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Mol Biochem Parasitol*, v. 124, p. 73-78, 2002.
- BROWN, A. J.; FISHER, D. A.; KOURANOVA, E. *et al.* Whole-rat conditional gene knockout via genome editing. *Nat Methods*, v. 10, p. 638-640, 2013.
- BUCKNER, F. S.; WILSON, A. J.; VAN VOORHIS, W. C. *Trypanosoma cruzi*: use of herpes simplex virus-thymidine kinase as a negative selectable marker. *Exp Parasitol*, v. 86, p. 171-180, 1997.
- BUSCAGLIA, C. A.; CAMPO, V. A.; FRASCH, A. C. *et al.* *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat Rev Microbiol*, v. 4, p. 229-236, 2006.
- CHAGAS, C. Nova Trypanozomíase Humana. *Memórias do Instituto Oswaldo Cruz*, v. p. 160-218, 1909.
- COOPER, R.; DE JESUS, A. R.; CROSS, G. A. Deletion of an immunodominant *Trypanosoma cruzi* surface glycoprotein disrupts flagellum-cell adhesion. *J Cell Biol*, v. 122, p. 149-156, 1993.
- COUGHLIN, B. C.; TEIXEIRA, S. M.; KIRCHHOFF, L. V. *et al.* Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. *J Biol Chem*, v. 275, p. 12051-12060, 2000.
- COURA, J. R. Chagas disease: control, elimination and eradication. Is it possible? *Mem Inst Oswaldo Cruz*, v. 108, p. 962-967, 2013.
- CROOKS, G. E.; HON, G.; CHANDONIA, J. M. *et al.* WebLogo: a sequence logo generator. *Genome Res*, v. 14, p. 1188-1190, 2004.
- CRUZ, M. C.; SOUZA-MELO, N.; DA SILVA, C. V. *et al.* *Trypanosoma cruzi*: role of delta-amastin on extracellular amastigote cell invasion and differentiation. *PLoS One*, v. 7, p. e51804, 2012.
- CUEVAS, I. C.; CAZZULO, J. J.; SANCHEZ, D. O. gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect Immun*, v. 71, p. 5739-5749, 2003.
- DALLAGIOVANNA, B.; PLAZANET-MENUT, C.; OGATTA, S. F. *et al.* *Trypanosoma cruzi*: a gene family encoding chitin-binding-like proteins is posttranscriptionally regulated during metacyclogenesis. *Exp Parasitol*, v. 99, p. 7-16, 2001.

- DANIELS, J. P.; GULL, K.; WICKSTEAD, B. Cell biology of the trypanosome genome. *Microbiol Mol Biol Rev*, v. 74, p. 552-569, 2010.
- DAROCHA, W. D.; OTSU, K.; TEIXEIRA, S. M. *et al.* Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol*, v. 133, p. 175-186, 2004a.
- DAROCHA, W. D.; SILVA, R. A.; BARTHOLOMEU, D. C. *et al.* Expression of exogenous genes in *Trypanosoma cruzi*: improving vectors and electroporation protocols. *Parasitol Res*, v. 92, p. 113-120, 2004b.
- DE FREITAS, J. M.; AUGUSTO-PINTO, L.; PIMENTA, J. R. *et al.* Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathog*, v. 2, p. e24, 2006.
- DE GAUDENZI, J. G.; D'ORSO, I.; FRASCH, A. C. RNA recognition motif-type RNA-binding proteins in *Trypanosoma cruzi* form a family involved in the interaction with specific transcripts in vivo. *J Biol Chem*, v. 278, p. 18884-18894, 2003.
- DE GODOY, L. M.; MARCHINI, F. K.; PAVONI, D. P. *et al.* Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. *Proteomics*, v. 12, p. 2694-2703, 2012.
- DE SOUZA, E. M.; DA SILVA, P. B.; NEFERTITI, A. S. *et al.* Trypanocidal activity and selectivity in vitro of aromatic amidine compounds upon bloodstream and intracellular forms of *Trypanosoma cruzi*. *Exp Parasitol*, v. 127, p. 429-435, 2011.
- DE SOUZA, F. S.; RAMPAZZO RDE, C.; MANHAES, L. *et al.* Knockout of the gene encoding the kinetoplast-associated protein 3 (KAP3) in *Trypanosoma cruzi*: effect on kinetoplast organization, cell proliferation and differentiation. *Mol Biochem Parasitol*, v. 172, p. 90-98, 2010.
- DE SOUZA, W. Basic cell biology of *Trypanosoma cruzi*. *Curr Pharm Des*, v. 8, p. 269-285, 2002.
- DOS SANTOS, S. L.; FREITAS, L. M.; LOBO, F. P. *et al.* The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLoS Negl Trop Dis*, v. 6, p. e1779, 2012.
- EICKHOFF, C. S.; VASCONCELOS, J. R.; SULLIVAN, N. L. *et al.* Co-Administration of a Plasmid DNA Encoding IL-15 Improves Long-Term Protection of a Genetic Vaccine against *Trypanosoma cruzi*. *PLoS Negl Trop Dis*, v. 5, p. e983, 2011.
- EL-SAYED, N. M.; MYLER, P. J.; BARTHOLOMEU, D. C. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, v. 309, p. 409-415, 2005.
- ELIAS, M. C.; NARDELLI, S. C.; SCHENKMAN, S. Chromatin and nuclear organization in *Trypanosoma cruzi*. *Future Microbiol*, v. 4, p. 1065-1074, 2009.
- FERREIRA, E. R.; BONFIM-MELO, A.; MORTARA, R. A. *et al.* *Trypanosoma cruzi* extracellular amastigotes and host cell signaling: more pieces to the puzzle. *Front Immunol*, v. 3, p. 363, 2012.

- GAVERIAUX-RUFF, C.; KIEFFER, B. L. Conditional gene targeting in the mouse nervous system: Insights into brain function and diseases. *Pharmacol Ther*, v. 113, p. 619-634, 2007.
- GIORGI, M. E.; DE LEDERKREMER, R. M. Trans-sialidase and mucins of *Trypanosoma cruzi*: an important interplay for the parasite. *Carbohydrate Research*, v. 346, p. 1389-1393, 2011.
- GLUENZ, E.; TAYLOR, M. C.; KELLY, J. M. The *Trypanosoma cruzi* metacyclic-specific protein Met-III associates with the nucleolus and contains independent amino and carboxyl terminal targeting elements. *International Journal for Parasitology*, v. 37, p. 617-625, 2007.
- GOMEZ, C.; ESTHER RAMIREZ, M.; CALIXTO-GALVEZ, M. *et al.* Regulation of gene expression in protozoa parasites. *J Biomed Biotechnol*, v. 2010, p. 726045, 2010.
- GOUY, M.; GUINDON, S.; GASCUEL, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, v. 27, p. 221-224, 2010.
- GUERRA-SLOMPO, E. P.; PROBST, C. M.; PAVONI, D. P. *et al.* Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs. *Biochemical and Biophysical Research Communications*, v. 420, p. 302-307, 2012.
- GUEVARA, P.; DIAS, M.; ROJAS, A. *et al.* Expression of fluorescent genes in *Trypanosoma cruzi* and *Trypanosoma rangeli* (Kinetoplastida: Trypanosomatidae): its application to parasite-vector biology. *J Med Entomol*, v. 42, p. 48-56, 2005.
- GUNZL, A.; BRUDERER, T.; LAUFER, G. *et al.* RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell*, v. 2, p. 542-551, 2003.
- HAJDUK, S.; OCHSENREITER, T. RNA editing in kinetoplastids. *RNA Biol*, v. 7, p. 229-236, 2010.
- HEIDMANN, D.; LEHNER, C. F. Reduction of Cre recombinase toxicity in proliferating *Drosophila* cells by estrogen-dependent activity regulation. *Dev Genes Evol*, v. 211, p. 458-465, 2001.
- HIROKAWA, T.; BOON-CHIENG, S.; MITAKU, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, v. 14, p. 378-379, 1998.
- IVENS, A. C.; PEACOCK, C. S.; WORTHEY, E. A. *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, v. 309, p. 436-442, 2005.
- JACKSON, A. P. The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Mol Biol Evol*, v. 27, p. 33-45, 2010.
- JANBANDHU, V. C.; MOIK, D.; FASSLER, R. Cre recombinase induces DNA damage and tetraploidy in the absence of LoxP sites. *Cell Cycle*, v. 13, p. 462-470, 2014.
- JULLIEN, N.; SAMPIERI, F.; ENJALBERT, A. *et al.* Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res*, v. 31, p. e131, 2003.

- JULLIEN, N.; GODDARD, I.; SELMI-RUBY, S. *et al.* Conditional transgenesis using Dimerizable Cre (DiCre). *PLoS One*, v. 2, p. e1355, 2007.
- KAHN, S.; VAN VOORHIS, W. C.; EISEN, H. The major 85-kD surface antigen of the mammalian form of *Trypanosoma cruzi* is encoded by a large heterogeneous family of simultaneously expressed genes. *J Exp Med*, v. 172, p. 589-597, 1990.
- KAHN, S.; COLBERT, T. G.; WALLACE, J. C. *et al.* The major 85-kDa surface antigen of the mammalian-stage forms of *Trypanosoma cruzi* is a family of sialidases. *Proc Natl Acad Sci U S A*, v. 88, p. 4481-4485, 1991.
- KANGUSSU-MARCOLINO, M. M.; DE PAIVA, R. M.; ARAUJO, P. R. *et al.* Distinct genomic organization, mRNA expression and cellular localization of members of two amastin sub-families present in *Trypanosoma cruzi*. *BMC Microbiol*, v. 13, p. 10, 2013.
- KELLY, J. M.; WARD, H. M.; MILES, M. A. *et al.* A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania*. *Nucleic Acids Res*, v. 20, p. 3963-3969, 1992.
- KIM, H. S.; LI, Z.; BOOTHROYD, C. *et al.* Strategies to construct null and conditional null *Trypanosoma brucei* mutants using Cre-recombinase and loxP. *Mol Biochem Parasitol*, v. 191, p. 16-19, 2013.
- KOLEV, N. G.; FRANKLIN, J. B.; CARMÍ, S. *et al.* The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog*, v. 6, p. e1001090, 2010.
- KRIEGER, M. A.; GOLDENBERG, S. Representation of Differential Expression: A New Approach to Study Differential Gene Expression in Trypanosomatids. *Parasitology Today*, v. 14, p. 163-166, 1998.
- KULKARNI, M. M.; OLSON, C. L.; ENGMAN, D. M. *et al.* *Trypanosoma cruzi* GP63 proteins undergo stage-specific differential posttranslational modification and are important for host cell infection. *Infect Immun*, v. 77, p. 2193-2200, 2009.
- LABAN, A.; WIRTH, D. F. Transfection of *Leishmania enriettii* and expression of chloramphenicol acetyltransferase gene. *Proc Natl Acad Sci U S A*, v. 86, p. 9119-9123, 1989.
- LARKIN, M. A.; BLACKSHIELDS, G.; BROWN, N. P. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics*, v. 23, p. 2947-2948, 2007.
- LAVERRIERE, M.; CAZZULO, J. J.; ALVAREZ, V. E. Antagonic activities of *Trypanosoma cruzi* metacaspases affect the balance between cell proliferation, death and differentiation. *Cell Death Differ*, v. 19, p. 1358-1369, 2012.
- LE, Y.; SAUER, B. Conditional gene knockout using Cre recombinase. *Mol Biotechnol*, v. 17, p. 269-275, 2001.
- LEE, M. G.; VAN DER PLOEG, L. H. Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol*, v. 51, p. 463-489, 1997.
- LI, Z. H.; DE GAUDENZI, J. G.; ALVAREZ, V. E. *et al.* A 43-nucleotide U-rich element in 3'-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes. *J Biol Chem*, v. 287, p. 19058-19069, 2012.

- LINSS, J.; GOLDENBERG, S.; URBINA, J. A. *et al.* Cloning and characterization of the gene encoding ATP-dependent phospho-enol-pyruvate carboxykinase in *Trypanosoma cruzi*: comparison of primary and predicted secondary structure with host GTP-dependent enzyme. *Gene*, v. 136, p. 69-77, 1993.
- LOW, H. P.; TARLETON, R. L. Molecular cloning of the gene encoding the 83 kDa amastigote surface protein and its identification as a member of the *Trypanosoma cruzi* sialidase superfamily. *Mol Biochem Parasitol*, v. 88, p. 137-149, 1997.
- MA, Y. F.; WEISS, L. M.; HUANG, H. A method for rapid regulation of protein expression in *Trypanosoma cruzi*. *Int J Parasitol*, v. 42, p. 33-37, 2012.
- MACRAE, J. I.; OBADO, S. O.; TURNOCK, D. C. *et al.* The suppression of galactose metabolism in *Trypanosoma cruzi* epimastigotes causes changes in cell surface molecular architecture and cell morphology. *Mol Biochem Parasitol*, v. 147, p. 126-136, 2006.
- MADEIRA DA SILVA, L.; OWENS, K. L.; MURTA, S. M. F. *et al.* Regulated expression of the *Leishmania* major surface virulence factor lipophosphoglycan using conditionally destabilized fusion proteins. *Proceedings of the National Academy of Sciences*, v. 106, p. 7583-7588, 2009.
- MANNING-CELA, R.; GONZALEZ, A.; SWINDLE, J. Alternative splicing of LYT1 transcripts in *Trypanosoma cruzi*. *Infect Immun*, v. 70, p. 4726-4728, 2002.
- MARTINEZ-CALVILLO, S.; LOPEZ, I.; HERNANDEZ, R. pRIBOTEX expression vector: a pTEX derivative for a rapid selection of *Trypanosoma cruzi* transfectants. *Gene*, v. 199, p. 71-76, 1997.
- MARTINEZ-CALVILLO, S.; NGUYEN, D.; STUART, K. *et al.* Transcription initiation and termination on *Leishmania* major chromosome 3. *Eukaryot Cell*, v. 3, p. 506-517, 2004.
- MARTINEZ-CALVILLO, S.; VIZUET-DE-RUEDA, J. C.; FLORENCIO-MARTINEZ, L. E. *et al.* Gene expression in trypanosomatid parasites. *J Biomed Biotechnol*, v. 2010, p. 525241, 2010.
- MARTINS-MELO, F. R.; RAMOS JUNIOR, A. N.; ALENCAR, C. H. *et al.* Multiple causes of death related to Chagas' disease in Brazil, 1999 to 2007. *Rev Soc Bras Med Trop*, v. 45, p. 591-596, 2012.
- MCNICOLL, F.; MULLER, M.; CLOUTIER, S. *et al.* Distinct 3'-untranslated region elements regulate stage-specific mRNA accumulation and translation in *Leishmania*. *J Biol Chem*, v. 280, p. 35238-35246, 2005.
- MENDES, T. A.; REIS CUNHA, J. L.; DE ALMEIDA LOURDES, R. *et al.* Identification of strain-specific B-cell epitopes in *Trypanosoma cruzi* using genome-scale epitope prediction and high-throughput immunoscreening with peptide arrays. *PLoS Negl Trop Dis*, v. 7, p. e2524, 2013.
- MILLS, A. A.; BRADLEY, A. From mouse to man: generating megabase chromosome rearrangements. *Trends Genet*, v. 17, p. 331-339, 2001.
- MINNING, T. A.; WEATHERLY, D. B.; ATWOOD, J., 3RD *et al.* The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genomics*, v. 10, p. 370, 2009.

- MYLER, P. J.; AUDLEMAN, L.; DEVOS, T. *et al.* Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A*, v. 96, p. 2902-2906, 1999.
- NILSSON, D.; GUNASEKERA, K.; MANI, J. *et al.* Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS Pathog*, v. 6, p. e1001037, 2010.
- NUNES, M. C.; DONES, W.; MORILLO, C. A. *et al.* Chagas disease: an overview of clinical and epidemiological aspects. *J Am Coll Cardiol*, v. 62, p. 767-776, 2013.
- OLIVAS-RUBIO, M.; HERNANDEZ-MARTINEZ, S.; TALAMAS-ROHANA, P. *et al.* cDNA cloning and partial characterization of amastigote specific surface protein from Trypanosoma cruzi. *Infect Genet Evol*, v. 9, p. 1083-1091, 2009.
- PAVIA, P. X.; THOMAS, M. C.; LOPEZ, M. C. *et al.* Molecular characterization of the short interspersed repetitive element SIRE in the six discrete typing units (DTUs) of Trypanosoma cruzi. *Exp Parasitol*, v. 132, p. 144-150, 2012.
- PEREIRA, C. A.; ALONSO, G. D.; IVALDI, S. *et al.* Arginine kinase overexpression improves Trypanosoma cruzi survival capability. *Febs Letters*, v. 554, p. 201-205, 2003.
- PIACENZA, L.; IRIGOIN, F.; ALVAREZ, M. N. *et al.* Mitochondrial superoxide radicals mediate programmed cell death in Trypanosoma cruzi: cytoprotective action of mitochondrial iron superoxide dismutase overexpression. *Biochem J*, v. 403, p. 323-334, 2007.
- PIRES, S. F.; DAROCHA, W. D.; FREITAS, J. M. *et al.* Cell culture and animal infection with distinct Trypanosoma cruzi strains expressing red and green fluorescent proteins. *Int J Parasitol*, v. 38, p. 289-297, 2008.
- REAL, F.; VIDAL, R. O.; CARAZZOLLE, M. F. *et al.* The genome sequence of Leishmania (Leishmania) amazonensis: functional annotation and extended analysis of gene models. *DNA Res*, v. 20, p. 567-581, 2013.
- ROCHETTE, A.; MCNICOLL, F.; GIRARD, J. *et al.* Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in Leishmania spp. *Mol Biochem Parasitol*, v. 140, p. 205-220, 2005.
- ROCHETTE, A.; RAYMOND, F.; UBEDA, J. M. *et al.* Genome-wide gene expression profiling analysis of Leishmania major and Leishmania infantum developmental stages reveals substantial differences between the two species. *BMC Genomics*, v. 9, p. 255, 2008.
- ROCHETTE, A.; RAYMOND, F.; CORBEIL, J. *et al.* Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of Leishmania infantum. *Mol Biochem Parasitol*, v. 165, p. 32-47, 2009.
- SANTOS, M. A.; GARG, N.; TARLETON, R. L. The identification and molecular characterization of Trypanosoma cruzi amastigote surface protein-1, a member of the trans-sialidase gene super-family. *Mol Biochem Parasitol*, v. 86, p. 1-11, 1997.
- SARKAR, I.; HAUBER, I.; HAUBER, J. *et al.* HIV-1 proviral DNA excision using an evolved recombinase. *Science*, v. 316, p. 1912-1915, 2007.

- SAUER, B. Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol*, v. 7, p. 2087-2096, 1987.
- SAUER, B.; HENDERSON, N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A*, v. 85, p. 5166-5170, 1988.
- SCAHILL, M. D.; PASTAR, I.; CROSS, G. A. CRE recombinase-based positive-negative selection systems for genetic manipulation in *Trypanosoma brucei*. *Mol Biochem Parasitol*, v. 157, p. 73-82, 2008.
- SCHMUNIS, G. Status of and cost of Chagas disease worldwide. *Lancet Infect Dis*, v. 13, p. 283-284, 2013.
- SCHOIJET, A. C.; MIRANDA, K.; MEDEIROS, L. C. *et al.* Defining the role of a FYVE domain in the localization and activity of a cAMP phosphodiesterase implicated in osmoregulation in *Trypanosoma cruzi*. *Mol Microbiol*, v. 79, p. 50-62, 2011.
- SHER, A.; SNARY, D. Specific inhibition of the morphogenesis of *Trypanosoma cruzi* by a monoclonal antibody. *Nature*, v. 300, p. 639-640, 1982.
- SIEGEL, T. N.; HEKSTRA, D. R.; WANG, X. *et al.* Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res*, v. 38, p. 4946-4957, 2010.
- SOUTHERN, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, v. 98, p. 503-517, 1975.
- SOUZA, R. T.; LIMA, F. M.; BARROS, R. M. *et al.* Genome size, karyotype polymorphism and chromosomal evolution in *Trypanosoma cruzi*. *PLoS One*, v. 6, p. e23042, 2011.
- STOBER, C. B.; LANGE, U. G.; ROBERTS, M. T. *et al.* From genome to vaccines for leishmaniasis: screening 100 novel vaccine candidates against murine *Leishmania major* infection. *Vaccine*, v. 24, p. 2602-2616, 2006.
- STRASEN, J.; WILLIAMS, T.; ERTL, G. *et al.* Epidemiology of Chagas disease in Europe: many calculations, little knowledge. *Clin Res Cardiol*, v. 103, p. 1-10, 2014.
- TAYLOR, M. C.; KELLY, J. M. pTcINDEX: a stable tetracycline-regulated expression vector for *Trypanosoma cruzi*. *BMC Biotechnol*, v. 6, p. 32, 2006.
- TEIXEIRA, S. M.; RUSSELL, D. G.; KIRCHHOFF, L. V. *et al.* A differentially expressed gene family encoding "amastin," a surface protein of *Trypanosoma cruzi* amastigotes. *J Biol Chem*, v. 269, p. 20509-20516, 1994.
- TEIXEIRA, S. M.; DAROCHA, W. D. Control of gene expression and genetic manipulation in the Trypanosomatidae. *Genet Mol Res*, v. 2, p. 148-158, 2003.
- TEIXEIRA, S. M.; DE PAIVA, R. M.; KANGUSSU-MARCOLINO, M. M. *et al.* Trypanosomatid comparative genomics: Contributions to the study of parasite biology and different parasitic diseases. *Genet Mol Biol*, v. 35, p. 1-17, 2012.
- TRITRYP DB: Kinetoplastid genomic resources Database. [<http://triTrypdb.org/common/downloads/release-4.1/Tcruzi/fasta/TriTrypDB>].

- TYLER, K. M.; ENGMAN, D. M. The life cycle of *Trypanosoma cruzi* revisited. *Int J Parasitol*, v. 31, p. 472-481, 2001.
- VAZQUEZ, M. P.; LEVIN, M. J. Functional analysis of the intergenic regions of TcP2beta gene loci allowed the construction of an improved *Trypanosoma cruzi* expression vector. *Gene*, v. 239, p. 217-225, 1999.
- VERBISCK, N. V.; DA-SILVA, S.; MORTARA, R. A. *Trypanosoma cruzi*: amastigote polymorphism defined by monoclonal antibodies. *Braz J Med Biol Res*, v. 31, p. 1583-1591, 1998.
- VERZIJLBERGEN, K. F.; FABER, A. W.; STULEMEIJER, I. J. *et al.* Multiple histone modifications in euchromatin promote heterochromatin formation by redundant mechanisms in *Saccharomyces cerevisiae*. *BMC Mol Biol*, v. 10, p. 76, 2009.
- WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics*, v. 10, p. 255, 2009.
- WHO. (2010) Report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis, pp 22-26 World Health Organization, Geneva.
- WILSON, T. J.; KOLA, I. The LoxP/CRE system and genome modification. *Methods Mol Biol*, v. 158, p. 83-94, 2001.
- WIRTZ, E.; LEAL, S.; OCHATT, C. *et al.* A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol*, v. 99, p. 89-101, 1999.
- XU, D.; BRANDAN, C. P.; BASOMBRIIO, M. A. *et al.* Evaluation of high efficiency gene knockout strategies for *Trypanosoma cruzi*. *BMC Microbiol*, v. 9, p. 90, 2009.
- YAMADA-OGATTA, S. F.; MOTTA, M. C.; TOMA, H. K. *et al.* *Trypanosoma cruzi*: cloning and characterization of two genes whose expression is up-regulated in metacyclic trypomastigotes. *Acta Trop*, v. 90, p. 171-179, 2004.
- ZINGALES, B.; ANDRADE, S. G.; BRIONES, M. R. *et al.* A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz*, v. 104, p. 1051-1054, 2009.

ANEXO 1



Trypanosomatid comparative genomics: Contributions to the study of parasite biology and different parasitic diseases

Santuza M. Teixeira¹, Rita Márcia Cardoso de Paiva¹, Monica M. Kangussu-Marcolino²
and Wanderson D. DaRocha²

¹*Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais,
Belo Horizonte, MG, Brazil.*

²*Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, PR, Brazil.*

Abstract

In 2005, draft sequences of the genomes of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*, also known as the Tri-Tryp genomes, were published. These protozoan parasites are the causative agents of three distinct insect-borne diseases, namely sleeping sickness, Chagas disease and leishmaniasis, all with a worldwide distribution. Despite the large estimated evolutionary distance among them, a conserved core of ~6,200 trypanosomatid genes was found among the Tri-Tryp genomes. Extensive analysis of these genomic sequences has greatly increased our understanding of the biology of these parasites and their host-parasite interactions. In this article, we review the recent advances in the comparative genomics of these three species. This analysis also includes data on additional sequences derived from other trypanosomatid species, as well as recent data on gene expression and functional genomics. In addition to facilitating the identification of key parasite molecules that may provide a better understanding of these complex diseases, genome studies offer a rich source of new information that can be used to define potential new drug targets and vaccine candidates for controlling these parasitic infections.

Key words: *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, genome, RNAseq.

Received: August 8, 2011; Accepted: October 18, 2011.

Tri-Tryp Diseases and The Tri-Tryp Genomes

Trypanosoma brucei, *Trypanosoma cruzi* and *Leishmania major* are unicellular protozoa of considerable medical importance since they are the etiologic agents of sleeping sickness (African trypanosomiasis), Chagas disease (American trypanosomiasis) and leishmaniasis, respectively. The geographic range of these parasites is determined by their insect vectors: in the case of sleeping sickness, a blood sucking fly of the genus *Glossina*, also known as the tsetse fly, for Chagas disease, a reduviid bug known as the “kissing bug” and for leishmaniasis, a phlebotomine sandfly. While sleeping sickness occurs in sub-Saharan Africa, Chagas disease is prevalent in Latin America. Leishmaniasis is considered to be endemic in 88 countries, 72 of which are developing countries in Asia, South America and Africa. Together, these three parasitic diseases represent a huge burden since approximately 0.5 million people are infected with *T. brucei*, 10 million with *T. cruzi* and an estimated 12 million with different species of *Leishmania*. In addition to the two human-infective subspe-

cies, *T. brucei gambiense* and *T. b. rhodesiense*, other species and subspecies of African trypanosomes cause the disease known as nagana in domestic animals, imposing a further economic burden on several African countries. Different forms of leishmaniasis are caused by at least 20 leishmanial species: cutaneous leishmaniasis, with an estimated 1.5 million cases, and visceral leishmaniasis, with about 500,000 new cases annually, are the most common. Although control of the arthropod vectors of these diseases is an achievable goal and has been successful against the *T. cruzi* vector in parts of Latin America, the alarming resurgence of sleeping sickness in Africa and of leishmaniasis in parts of Asia and Latin America is a constant reminder of the need for better forms of chemotherapy and prevention of these diseases. More detailed information on African and American trypanosomiasis and the different forms of leishmaniasis, including vector distribution, disease control and treatment protocols can be found at <http://apps.who.int/tdr/>.

Trypanosoma brucei, *T. cruzi* and *Leishmania* spp. are hemoflagellates of the family Trypanosomatidae (order Kinetoplastida) that is characterized by the presence of a single flagellum and one mitochondrion containing a unique organelle known as the kinetoplast which contains the

Send correspondence to Wanderson Duarte DaRocha. Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Caixa Postal 19046, 81531-990 Curitiba, PR, Brazil. E-mail: darocha@ufpr.br.

mitochondrial DNA (Simpson *et al.*, 2006). Each parasite has a complex life cycle that involves humans as one of their various hosts (Figure 1). As some of the earliest divergent members of the Eukaryotae (Haag *et al.*, 1998), these parasites have peculiar aspects of gene expression, including polycistronic transcription of most of their genomes (Martínez-Calvillo *et al.*, 2010), RNA polymerase I-mediated transcription of protein-coding genes (Gunzl *et al.*, 2003), RNA trans-splicing to generate mature, capped mRNAs (LeBowitz *et al.*, 1993) and extensive RNA editing to generate functional mRNAs transcribed from mitochondrial genes (Hajduk *et al.*, 1993). Apart from their medical relevance, these peculiar characteristics make these parasites very interesting models for studying genome evolution and other aspects of genome function. On the other hand, the early evolutionary divergence of these organisms has resulted in biochemical characteristics that are not common in higher eukaryotes, such as enzymes related to antioxidant metabolism (Olin-Sandoval *et al.*,

2010) as well as sterol and glycosylphosphatidylinositol (GPI) biosynthesis (Lepesheva *et al.*, 2011; Koeller and Heise, 2011) that have been exploited as promising drug targets.

Genome sequencing of Tri-Tryp parasites began in the early 90s with the analyses of 518 expressed sequence tags (ESTs) generated from mRNA isolated from blood-stream forms of *T. b. rhodesiense* (El-Sayed *et al.*, 1995). Shortly thereafter, a comparison between EST and genomic sequences showed that sequencing random DNA fragments was as efficient as EST analyses for discovering new genes in the African trypanosome (El-Sayed and Donelson, 1997). In 1996, an EST analysis of cDNA libraries constructed with mRNA from *L. major* promastigotes was published (Levick *et al.*, 1996), and the first EST analysis of *T. cruzi* epimastigote forms was published in 1997 (Brandão *et al.*, 1997). During this period, pulsed-field gel electrophoretic analysis of chromosomes and the sequencing of large DNA fragments from cosmid, bacterial artificial chromosome and yeast artificial chromosome libraries were also undertaken to generate physical maps of Tri-Tryp genomes (Blackwell and Melville, 1999). In 1999, the sequence of a 257-kilobase region spanning almost the entire chromosome 1 of *L. major* revealed the unusual distribution of protein-coding genes that was later found to be characteristic of all Tri-Tryp genomes. The complete sequence of *L. major* chromosome 1 revealed 79 protein-coding genes, with the first 29 genes all encoded on one DNA strand and the remaining 50 genes encoded on the opposite strand (Myler *et al.*, 1999).

The Tri-Tryp gene organization is reminiscent of bacterial operons, with protein coding genes densely packed within directional clusters in one strand separated by strand switch regions (*i.e.*, changes in the coding strand) (Figure 2). Experimental evidence suggests that transcription initiates bi-directionally between two divergent gene clusters (Martínez-Calvillo *et al.*, 2003, 2004) to produce polycistronic pre-mRNAs that are subsequently processed. Remarkably, with the exception of the spliced leader (SL) promoter, no promoter is recognized by RNA polymerase II and only a few transcription factors have been identified (Cribb and Serra, 2009; Cribb *et al.*, 2010). Even more surprisingly, although orthologs of all conserved components of the RNA polymerase II complex were identified in the Tri-Tryp genome (Ivens *et al.*, 2005), the transcription of some trypanosomatid genes such as VSG (Variant Surface Glycoprotein) and the procyclin genes of *T. brucei*, as well as several exogenous genes transfected into *T. cruzi*, are mediated by RNA polymerase I (Gunzl *et al.*, 2003). Once the polycistronic pre-mRNA is produced, two coupled reactions (trans-splicing and poly-adenylation) result in mature monocistronic transcripts.

Trans-splicing means that every mature mRNA has an identical capped sequence of 39 nucleotides, known as the spliced leader (SL), at the 5' end (Liang *et al.*, 2003).

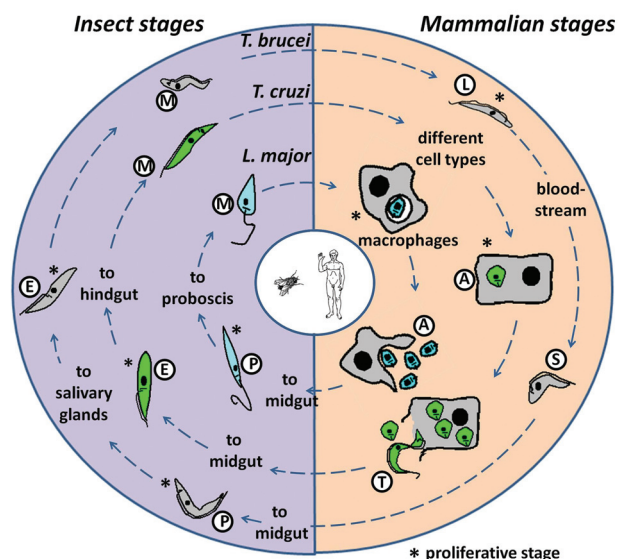


Figure 1 - The Tri-Tryp life cycles. Representation of the life cycles of *Leishmania major*, *Trypanosoma cruzi* and *T. brucei*, the etiological agents of leishmaniasis, Chagas disease and sleeping sickness, respectively, are shown, with the parasitic forms that are present in the insect vectors and the mammalian hosts. *Leishmania major* proliferates as promastigotes (P) in the sand fly midgut. The parasite is transmitted during bites by this fly and invades mammalian macrophages in the metacyclic promastigote (M) form. Inside the cell, the M form is converted into amastigotes (A) and divides before being released during cell lysis. *Trypanosoma cruzi* replicates as epimastigotes (E) in the reduviid bug midgut and develops into infective metacyclic trypomastigotes that are excreted in the feces (M) and invade different cell types when in contact with the mammalian host. After differentiation into proliferative amastigotes (A), these are transformed into bloodstream trypomastigotes (T) that cause cell lysis and invade new cells. *Trypanosoma brucei* differentiates from procyclic (P) to epimastigote (E) proliferative forms in the tsetse fly before being transformed into infective, metacyclic forms (M) in the salivary glands. After being injected into the host during a blood meal, M forms differentiate into long slender forms (L) that proliferate in the bloodstream and can reach the central nervous system. After increase of parasite numbers these last forms are replaced by non-proliferative stumpy forms (S).

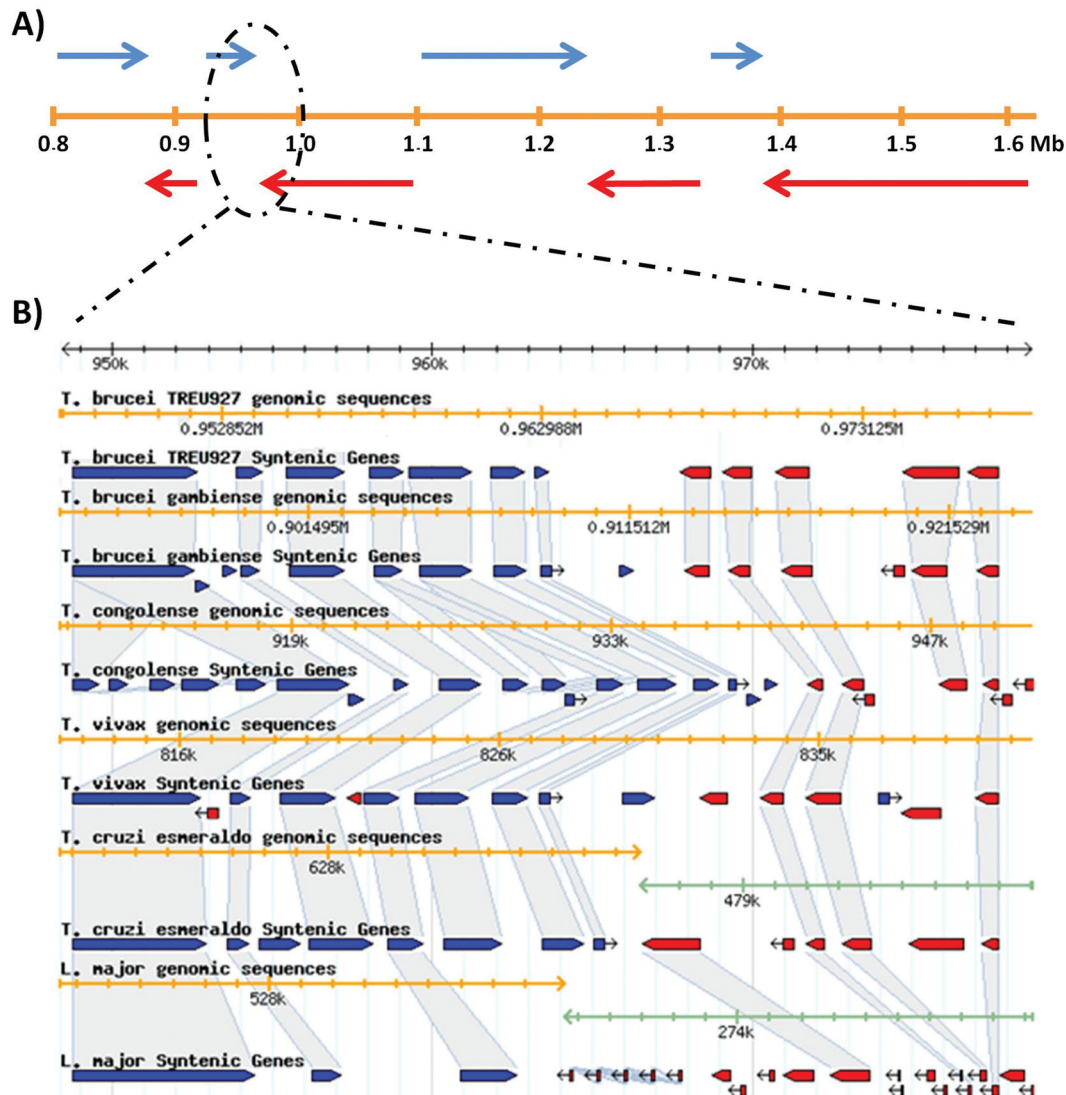


Figure 2 - Gene organization in the Tri-Tryp genome. Panel A shows the gene distribution in a 0.8 Mb region of *T. brucei* chromosome V with eight large polycistronic transcription units (blue arrows: plus strand encoded open reading frames or ORFs; red arrows: minus strand encoded ORFs). In panel B, a genomic region at around 960 kb is magnified to show the gene synteny in the genomes of various trypanosomatids (blue and red boxes correspond to + and - strand-encoded ORFs, respectively). The orange line in both panels corresponds to the chromosome position. Sequence information used to draw panel A and the graphic representation in panel B were obtained from the Tri-Tryp database (Aslett *et al.*, 2010).

Whilst no sequence consensus for polyadenylation or SL addition has been found, several studies have demonstrated that polypyrimidine-rich tracts located within intergenic regions guide SL addition and poly-adenylation, resulting in mature mRNAs (LeBowitz *et al.*, 1993) (Figure 3). Intergenic sequences involved in the processing of *T. cruzi*, *T. brucei* and *Leishmania* mRNA have been thoroughly investigated by comparing mRNA with genomic sequences, initially using EST databases (Benz *et al.*, 2005; Campos *et al.*, 2008; Smith *et al.*, 2008) and, more recently, using high-throughput RNA-sequencing (RNAseq) (Siegel *et al.*, 2010; Kolev *et al.*, 2010; Nilsson *et al.*, 2010). In addition to providing valuable information on the mechanisms of gene expression in these organisms, these analyses also

yielded data that allowed the optimization of transfection vectors used to express foreign genes and genetic manipulation in trypanosomatids.

Comparative genomic analyses using the Tri-Tryp sequences have already provided interesting insights into the genetic and evolutionary bases of the distinct and shared lifestyles of these parasites. Probably the most striking finding is that the three genomes display high levels of synteny and share a conserved set of ~6,200 genes, 94% of which are arranged in syntenic directional gene clusters (El-Sayed *et al.*, 2005a). Alignment of the deduced protein sequences of the majority of the clusters of orthologous genes across the three organisms reveals an average 57% identity between *T. cruzi* and *T. brucei* and 44% identity be-

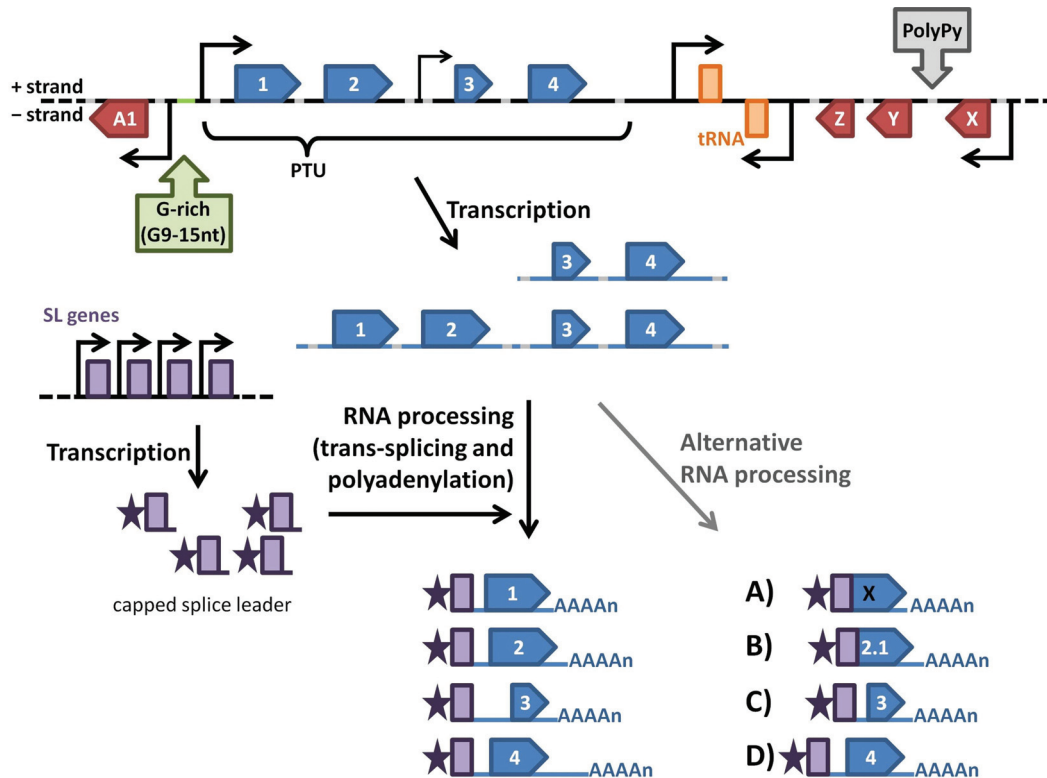


Figure 3 - Gene expression in trypanosomatids. Large clusters of unrelated genes (arrow boxes) are organized as polycistronic transcription units (PTUs) that are separated by divergent or convergent strand-switch regions. RNA Pol II transcription start sites (TSS) are usually located upstream of the first gene of the PTU (Martínez-Cavillo *et al.*, 2004) or can be located as an internal TSS (Kolev *et al.*, 2010). At the TSS (large bent arrow), the histone variants H2AZ and H2BV (Siegel *et al.*, 2009), modified histones [K9/K14 acetylated and K4 tri-methylated histone (Respuela *et al.*, 2008; Thomas *et al.*, 2009; Wright *et al.*, 2010) and K10 acetylated histone H4 (Siegel *et al.*, 2009)], bromodomain factor BDF3 (Siegel *et al.*, 2009) and transcription factors TRF4 and SNAP50 (Thomas *et al.*, 2009) are frequently associated, with a few of these chromatin modifications also detected at internal TSS (small bent arrow) (Siegel *et al.*, 2009; Wright *et al.*, 2010). The polycistronic RNAs (pre-mRNAs) are individualized in monocistronic mRNAs after the addition of a capped splice leader RNA through a trans-splicing reaction coupled to polyadenylation. These processing reactions are guided by polypyrimidine tracts (PolyPy) that are present in every intergenic region. Mature mRNAs are exported to the cytoplasm where their stability and translation efficiencies are largely dependent on cis-acting elements present in their untranslated region (UTR) (Araujo *et al.*, 2011). Transcriptomic analyses also showed that polycistronic pre-mRNAs can suffer alternative RNA processing that may result in changes in the initiator AUG, thereby altering protein translation (A), targeting and/or function (B). Alternative splicing and poly-adenylation can also result in the inclusion/exclusion of regulatory elements present in the 5' UTRs (C) or 3' UTRs (D), thereby altering gene expression (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010).

tween *T. cruzi* and *L. major* that reflected the expected phylogenetic relationships (Lukes *et al.*, 1997; Haag *et al.*, 1998; Stevens *et al.*, 1999; Wright *et al.*, 1999). The majority of species-specific genes occurs on non-syntenic chromosomes and consists of members of large surface antigen families. Structural RNAs, retroelements and gene family expansion are also often associated with breaks in the conservation of gene synteny (El-Sayed *et al.*, 2005a). Multi-gene family expansions are generally species-specific and most pronounced in the *T. cruzi* genome. As discussed below, a number of *T. cruzi* multi-gene families encode surface proteins, such as trans-sialidases, mucin-associated surface proteins (MASP) and mucins TcMUC and GP63 that likely play important roles in host-parasite interactions (Di Noia *et al.*, 1995; Vargas *et al.*, 2004; Baida *et al.*, 2006; Bartholomeu *et al.*, 2009). Based on their location in regions of synteny breaks these arrays may be subject to extensive rearrangements during the parasite's evolution and

are thus directly associated with the specificities of each of the three parasitic diseases.

The Genetic Diversity of *T. cruzi* and the Genomes of Different Parasite Strains

Chagas disease, caused by *T. cruzi*, is endemic in more than 20 Latin American countries, where an estimated 10 million people are infected and the "domiciliation" of the triatomines exposes at least 90 million individuals to the risk of infection. With no vaccine or effective drug treatment available, the main strategy for control must rely on the prevention of transmission by the insect vectors and blood transfusions. The parasite proliferates in the midgut of several species of a triatomid hematophagous vector. After reaching the insect's hindgut, epimastigote forms differentiate into non-dividing, infective metacyclic trypomastigotes that are excreted in the insect's feces. Trypomastigotes can infect a mammalian host by passing through

mucous membranes or skin lesions during feeding by the insect. Once inside the mammalian host, trypomastigotes invade different types of cells where they transform into proliferative intracellular amastigotes. After a number of cell divisions in the host cell cytoplasm, amastigotes differentiate into trypomastigotes that are released into the bloodstream after host cell rupture and, after being taken up by an insect during a blood meal, they start a new cycle (Brenner, 1973) (Figure 1). The highly heterogeneous *T. cruzi* population consists of a large number of strains with distinct characteristics related to morphology, growth rate, parasitemia curves, virulence, pathogenicity, drug sensitivity, antigenic profile, metacyclogenesis and tissue tropism (Buscaglia and Di Noia, 2003).

Despite the broad genetic diversity observed among different strains and isolates, early studies based on different genotyping strategies identified two major lineages in the parasite population, named *T. cruzi* I and *T. cruzi* II (Souto *et al.*, 1996; Momen 1999). These divergent lineages occupy distinct ecological environments, namely, the sylvatic cycle (*T. cruzi* I) and the domestic cycle (*T. cruzi* II) of Chagas disease (Zingales *et al.*, 1998), as well as distinct sylvatic host associations (Buscaglia and Di Noia, 2003). Further analyses led some authors to propose the sub-division of *T. cruzi* II into five sub-groups: *T. cruzi* IIa, IIb, IIc, IId and IIE (Brisse *et al.*, 2000). Phylogenetic analyses of the *T. cruzi* strains became more confusing when additional data indicated the existence of not just two, but three major groups in the *T. cruzi* population, in addition to hybrid strains (Miles *et al.*, 1978; Augusto-Pinto *et al.*, 2003; de Freitas *et al.*, 2006). After intense debate, in 2009 an international consensus recognized the existence of six major strains, also known as discrete typing units (DTUs) I-VI (Zingales *et al.*, 2009) (Table 1). Since Chagas disease spawns a variety of clinical forms, these studies are highly relevant: understanding the genetic variation among strains can potentially explain differences in disease pathogenesis, host preferences and, most importantly, provides essential information for the identification of new drug targets and good antigenic candidates for better diagnosis and vaccine development. For instance, *T. cruzi* II strains and the hybrid

strains belonging to *T. cruzi* V and VI are the predominant causes of human disease in South America (Zingales *et al.*, 2009), whereas *T. cruzi* I strains are more abundant among wild hosts and vectors. Although detailed analysis of the biological and molecular factors underlying *T. cruzi* population structure and the epidemiology of Chagas disease are beyond the scope of this review, one must keep in mind that the genetic variability found in the *T. cruzi* population is an essential aspect to be considered when analyzing this parasite's genome.

CL Brenner, a clone derived from a hybrid *T. cruzi* strain belonging to *T. cruzi* VI, was chosen as a reference strain for the initial *T. cruzi* genome project. The hybrid nature of the CL Brenner clone became clear only after the genome sequencing had begun, when analyses of nuclear and mitochondrial sequences showed that this strain resulted from a fusion event that had occurred between ancient genotypes corresponding to strains belonging to *T. cruzi* II and III groups (El-Sayed *et al.*, 2005a; de Freitas *et al.*, 2006). Prior to this knowledge, the choice of the clone CL Brenner, initially classified as a member of sub-group IIE, was based on five characteristics: (1) it was isolated from the domiciliary vector *Triatoma infestans*, (2) its pattern of infectivity in mice was very well known, (3) it had preferential tropism for heart and muscle cells, (4) it showed a clear acute phase in accidentally infected humans, and (5) it was susceptible to drugs used to treat Chagas disease (Zingales *et al.*, 1997). In addition, several genomic studies had previously used this strain for karyotype analyses (Branche *et al.*, 2006) and the generation of physical maps and ESTs from all three stages of the parasite life cycle (Cano *et al.*, 1995; Henriksson *et al.*, 1995; Brandão *et al.*, 1997; Verdun *et al.*, 1998; Porcel *et al.*, 2000; Cerqueira *et al.*, 2005).

The *T. cruzi* CL Brenner haploid genome, estimated to be 55 Mb, was sequenced using the WGS (whole genome shotgun) strategy. Because of its hybrid nature and the high level of allelic polymorphism, a 14X coverage, much higher than the usual 8-10X coverage, was required to distinguish the ambiguities derived from allelic variations from those produced by sequencing errors. In contrast to the other two Tri-Tryp genomes, the *T. cruzi* draft sequence (El-Sayed *et al.*, 2005b) was published as an assembly of 5,489 scaffolds built by 8,740 contigs. Four years later, based on synteny maps for the *T. brucei* chromosomes, Weatherly *et al.* (2009) assembled the *T. cruzi* contigs and scaffolds initially in 11 pairs of homologous “*T. brucei*-like” chromosomes and, ultimately, in 41 *T. cruzi* chromosomes. Since trypanosomatid chromosomes do not condense during mitosis and are therefore not visualized in metaphasic cells the predicted number of *T. cruzi* chromosomes was based on studies of pulsed-field gel electrophoresis (PFGE) analyses (Branche *et al.*, 2006), which turned out to be similar to the number of assembled chromosomes. As mentioned above, the genome organization in *T. cruzi* is largely syntenic with the other Tri-Tryp (*T. brucei* and *L.*

Table 1 - Classification of *T. cruzi* strains.

Current designation ^a	Equivalence to former classifications ^b	Examples of representative strains
<i>T. cruzi</i> I	<i>T. cruzi</i> I/DTU I	Sylvio X-10, Dm28c
<i>T. cruzi</i> II	<i>T. cruzi</i> II/DTU IIb	Esmeraldo, Y
<i>T. cruzi</i> III	<i>T. cruzi</i> III/DTU IIc	CM17
<i>T. cruzi</i> IV	DTU IIa	CanIII
<i>T. cruzi</i> V ^c	DTU IId	SO3
<i>T. cruzi</i> VI ^c	DTU IIE	CL Brenner

DTU = discrete typing unit. ^aZingales *et al.* (2009), ^bMomen (1999) (*T. cruzi* I and II classification), Brisse *et al.* (2000) (DTU I, IIa-e), de Freitas *et al.* (2006) (*T. cruzi* I, II and III), ^cHybrid strains.

major) genomes, with most species-specific genes, such as surface protein gene families, occurring in internal and subtelomeric regions of non-syntenic chromosome (El-Sayed *et al.*, 2005a).

Because of its hybrid nature, the CL Brener genome is represented by a redundant dataset since homologous regions displaying a high level of polymorphism were assembled separately, generating two set of contigs, each corresponding to one haplotype. To identify the two haplotypes, reads from the genome of the cloned Esmeraldo strain, a member of *T. cruzi* II, and representing one of the CL Brener parental strain (de Freitas *et al.*, 2006), were generated. Thus, in the annotation data of the CL Brener genome, the two haplotypes are referred to as “Esmeraldo-like” or “non-Esmeraldo-like” sequences (Aslett *et al.*, 2010).

The haploid CL Brener genome has an estimated 12,000 genes. As with the other Tri-Tryps, the *T. cruzi* genes are organized in long polycistronic clusters that are transcribed by RNA polymerase II and processed into monocistronic mRNAs that accumulate differentially during the various stages of the parasite life cycle. As indicated before, one of the main characteristics revealed by the complete sequence of the *T. cruzi* genome was the dramatic expansion of families encoding surface proteins (El-Sayed *et al.*, 2005a). Compared to *T. brucei* and *L. major*, *T. cruzi* has the largest set of multi-gene families, perhaps because of its unique capacity to invade and multiply within different types of host cells. Long terminal repeat (LTR) and non-LTR retroelements and other sub-telomeric also contribute to the large proportion of repetitive sequences (50% of the genome) in this genome. The largest protein gene family encodes a group of surface proteins known as trans-sialidases (TS), with 1,430 members. TSs are surface molecules identified as virulent factors of *T. cruzi* that are responsible for transferring sialic acid from host sialoglycoconjugates to the terminal β -galactose on *T. cruzi* mucins. Mucin-associated surface proteins (MASP) are the second largest *T. cruzi* gene family, with a total of 1,377 members. Although MASP sequences correspond to ~6% of the parasite diploid genome, they were only identified during annotation of the *T. cruzi* genome. MASPs are glycosylphosphatidylinositol (GPI)-anchored surface proteins that are preferentially expressed in trypomastigotes; these proteins are characterized by highly conserved N- and C-terminal domains and a strikingly variable and repetitive central region (Bartholomeu *et al.*, 2009). Together with the mucin and GP63 gene families, these four gene families account for ~17% of all protein-coding genes and are organized as dispersed clusters of tandem and interspersed repeats.

Other large families consist of the previously described RHS and DGF-1 genes whose functions are unknown and which, like the TS genes, occur mostly at sub-telomeric locations. Examples of other gene families with more than 10 members also present in the *T. cruzi* ge-

nome include glycosyltransferases, protein kinases and phosphatases, kinesins, amino acid transporters and helicases, in addition to several gene families encoding hypothetical proteins (El-Sayed *et al.*, 2005a). The collapse of nearly identical repeats in some gene families, such as the gene cluster encoding α - and β -tubulins, meant that not all copies of the family were included in the original genome assembly.

Arner *et al.* (2007) described an analysis of the total genomic repetitive content of protein coding sequences and concluded that 18% of all protein coding sequences existed in 14 or more copies. In addition to the need to evade the host immune system, the existence of highly repetitive gene families in the *T. cruzi* genome in which a large number of gene copies can lead to the enhanced expression of various proteins may help to overcome a major problem in this genome, namely, the lack of strong promoters capable of generating high levels of mRNA from single copy genes. It is also likely that many of the striking polymorphisms among *T. cruzi* isolates that are reflected in several epidemiological and pathological aspects of Chagas disease are partly attributable to variability within regions containing gene families. Whole genome comparisons of distinct *T. cruzi* lineages are beginning to improve our understanding of this question.

Soon after the CL Brener genome was completed several groups began sequencing the genome of representative strains of other major *T. cruzi* lineages. As indicated above, the hybrid nature of the CL Brener genome provided data for two genomes, with “Esmeraldo-like” and “non-Esmeraldo” contigs making it possible to distinguish information from *T. cruzi* II and III groups, respectively (see Table 1). Recently, Franzén *et al.* (2011) published a draft genome sequence of Sylvio X10, a strain belonging to *T. cruzi* I group, which is the predominant agent of Chagas disease in Central America and in the Amazon. Although rarely isolated from humans in endemic areas in southern countries of Latin America where most cases of Chagas disease with mega-syndromes occur, *T. cruzi* I strains are highly abundant among wild hosts and vectors (Zingales *et al.*, 1998; Buscaglia and Di Noia, 2003). Thus, the distinct ecological niches occupied by *T. cruzi* I and II strains, together with the fact these strains are highly divergent in terms of phylogenetic analysis, prompted Franzén *et al.* (2011) to sequence the genome of a representative of *T. cruzi* I group and to undertake a comparative analysis with the CL Brener genome.

In agreement with previous analyses, the Sylvio X10 genome was estimated to be ~44 Mb in size, *i.e.*, smaller than the CL Brener genome. Indeed, smaller genomes seems to be a general feature of *T. cruzi* I strains (Branche *et al.*, 2006; Franzén *et al.*, 2011). As expected, the architectures of the two genomes were very similar, with highly conserved syntenic regions corresponding to the gene-dense “core” of the coding regions organized for long

polycistronic transcription. As with the CL Brener genome, the presence of repetitive sequences meant that the Sylvio X10 genome was represented as fragmented contigs. The technical difficulties associated with the assembly of repetitive sequences meant that only about 49% of the generated Sylvio X10 sequence data was incorporated into contigs, leaving 710,109 reads that were not included in the assembly. Consequently, the draft genome of Sylvio X10 was assembled into 7,092 contigs, which is slightly less than the number of contigs reported for the draft genome of CL Brener. The alignment of these contigs to both CL Brener haplotypes showed that the mean nucleotide identity was greater between Sylvio X10 and non-Esmeraldo (98.2%) than between Sylvio X10 and Esmeraldo (97.5%). This finding agrees with previous phylogenetic analyses indicating that sequences from *T. cruzi* I strains are more closely related to *T. cruzi* III (represented by the non-Esmeraldo CL Brener haplotype) than to *T. cruzi* II (represented by the Esmeraldo-like haplotype) (Cerqueira *et al.*, 2008; Ruvalcaba-Trejo and Sturm, 2011).

In contrast to the hybrid CL Brener genome, for which the amount of heterozygosity in the core genome was estimated to be 5.5% (El-Sayed *et al.*, 2005a), the diploid Sylvio X10 genome was homozygous (< 0.08% heterozygosity). Most importantly, analysis of the core gene content of CL Brener and Sylvio X10 revealed six open reading frames that were missing in the Sylvio X10 genome. Besides these six genes, estimations based on total sequence reads indicated that several multicopy gene families, including DGF, mucin, MASP and GP63 contained substantially fewer genes in Sylvio X10 than in CL Brener. A 5.9 Mb size difference between the Sylvio X10/1 and CL Brener genomes largely reflected the expansion of these gene families. However, the extent to which these genomic variations are related to strain differences in host preference and the ability to cause Chagas disease remains to be determined.

The advent of next-generation sequencing technologies has ushered in a new era in comparative sequencing by allowing the exploration of a wide range of evolutionary and pathological questions within the *T. cruzi* lineage. Several groups have initiated sequencing analyses of additional *T. cruzi* isolates. A consortium of laboratories funded by the National Institutes of Health/National Institutes of Allergy and Infectious Diseases (NIAID) and the National Human Genome Research Institute (NHGRI) is sequencing *T. cruzi* strains representative of each one of the six main groups, such as Esmeraldo (*T. cruzi* II), 3869 (*T. cruzi* III), Can III (*T. cruzi* IV), NRcl3 (*T. cruzi* V) and Tula cl2 (*T. cruzi* VI) (N. El-Sayed, personal communication). Our laboratory has been involved in the sequencing of another *T. cruzi* I strain (Dm28c) and CL-14, a non-virulent strain that belongs to the *T. cruzi* VI group (S. Teixeira, unpublished). In contrast to CL Brener, BALB/c mice injected with CL-14 trypomastigotes showed no parasitemia but devel-

oped high resistance against a lethal challenge with virulent trypomastigotes from the CL Brener or Y strains (Lima *et al.*, 1995). Our goal in this work is to use comparative analyses of the CL Brener and CL-14 genomes to identify potential sequences that can restore the virulence of CL-14 and then test these in transfection protocols.

In addition to investigations of the nuclear genome, several studies have examined the mitochondrial genome of kinetoplastids which contains a mass of concatenated DNA known as kinetoplast DNA (kDNA) that is easily identified near the insertion of the flagellum (Brener, 1973). In *T. cruzi*, kDNA consists of a highly structured disk-shaped network of thousands of concatenated minicircles 0.5-10 kb in size and dozens of concatenated maxicircles 20-40 kb in size. Whereas minicircle sequences are present exclusively in kinetoplastids, maxicircles are the homologues of mtDNA molecules found in other eukaryotes (Lukes *et al.*, 1997). Following publication of the *T. cruzi* genome, Westenberger *et al.* (2006) described the complete sequences of maxicircle DNAs corresponding to groups *T. cruzi* II (from sequences of the Esmeraldo strain) and III (from CL Brener sequences). As with other trypanosomatid mitochondrial genes, sequence analyses showed that *T. cruzi* maxicircle DNA contained frameshift errors in most of its genes that were corrected at the RNA level by a complex U-insertion/deletion process known as RNA editing (Hajduk *et al.*, 1993). Key elements of this repair process include gRNAs (guide RNAs) which are encoded mainly by minicircles, although a few gRNA sequences are also present in maxicircles. The gRNAs hybridize to the 3' end of a target message and undertake direct U insertion and deletion by the so-called editosome machinery (Stuart and Panigrahi, 2002).

The complete sequences of the 25 kb *T. cruzi* maxicircles revealed 18 tightly clustered mitochondrial protein-coding genes and two rRNA genes that were syntenic with previously sequenced maxicircles of *T. brucei* and *Leishmania tarentolae*. Fifteen of the 18 protein-coding genes were edited. Outside the coding region, strain-specific repetitive regions and a variable region that was unique for each strain were identified (Westenberger *et al.*, 2006). More recently, comparative analyses of the mitochondrial genomes of *T. cruzi* I, II and III were reported after Ruvalcaba-Trejo and Sturm (2011) generated the sequence of the coding region of the maxicircle from Sylvio X10. In agreement with the nuclear genomic analysis, phylogenetic analysis of the maxicircle coding regions supported a close evolutionary relationship between *T. cruzi* I and III. Based on their mitochondrial DNA analyses, these authors proposed a model in which an ancestral strain belonging to *T. cruzi* I provided the maxicircle for the progeny of a TcI-TcII hybridization event that resulted in the generation of *T. cruzi* III and *T. cruzi* IV strains. A subsequent 'back-cross' hybridization between *T. cruzi* II and *T. cruzi* III strains resulted in the *T. cruzi* V and VI strains, such as

CL Brener, that carry the maxicircle from their *T. cruzi* III ancestor.

Comparative Genomics of Leishmania Species That Cause Distinct Forms of Leishmaniasis

Leishmania spp. are parasitic protozoa transmitted by the bites of phlebotomine sand flies that are endemic in tropical and subtropical regions worldwide. More than 20 species are responsible for a wide spectrum of diseases, known as leishmaniasis (Murray *et al.*, 2005). Parasites in this genus are classified into two subgenera according to the part of the sandfly gut where colonization and development occur: the subgenus *Leishmania* (*Leishmania*) consists of parasites with mid and foregut development, whereas the subgenus *Leishmania* (*Viannia*) consists of parasites that undergo hindgut development (Lainson *et al.*, 1977; Bates, 2007). Depending on the species of *Leishmania*, infection of humans may result in diverse clinical forms of leishmaniasis with symptoms ranging from self-healing cutaneous lesions (*L. major*/*L. tropical*/*L. mexicana*) to fatal visceral leishmaniasis (*L. donovani*/*L. infantum*/*L. chagasi*). Infection by *Leishmania* can also result in mucosal leishmaniasis (mainly caused by *L. braziliensis*) and diffuse cutaneous leishmaniasis (mainly caused by *L. amazonensis*/*L. guyanensis*/*L. aethiopica*) (Desjeux, 1996). In addition to the species of *Leishmania*, other factors such as the genetic variability of the human host may determine the disease tropism and clinical manifestations in leishmaniasis (Blackwell *et al.*, 2009; Sakthianandeswaren *et al.*, 2009).

The World Health Organization (WHO) estimates that there are over two million new cases of leishmaniasis each year, with more than 360 million people at risk of contracting this disease in 88 countries on five continents (Asia, Africa, Europe, North America and South America) (www.who.int/tdrdiseases/leish). As part of their life cycle, *Leishmania* spp. alternate between the alimentary tract of the sandfly vector, where they grow as extracellular flagellated promastigotes and differentiate into infective non-dividing metacyclic forms, and the phagolysosome of the vertebrate host macrophages, where they differentiate into aflagellated, replicative amastigotes (Figure 1). There is no effective vaccine against *Leishmania* and the available therapeutic arsenal is extremely limited (Mauel, 2002). Thus, completion of the genome sequences of several *Leishmania* species (Ivens *et al.*, 2005; Peacock *et al.*, 2007) represents a long awaited aspiration for groups involved in the discovery and development of new drugs and vaccine targets.

Leishmania major Friedlin was chosen as the *Leishmania* reference strain for the Tri-Tryp genome project. The *L. major* haploid genome (~32.8 Mb) is distributed among 36 relatively small chromosomes ranging from 0.28 to 2.8 Mb in size (Wincker, 1996) and was sequenced

after shotgun cloning of large DNA fragments derived from chromosomal bands separated in agarose gels. Prior to publication of the Tri-Tryp genome sequence, the complete sequences of chromosomes 1 and 3 from *L. major* were published (Myler *et al.*, 1999; Worthey *et al.*, 2003) and an optical map of the entire genome was generated (Zhou *et al.*, 2004). In 2007, the complete genomes of two other *Leishmania* species, *L. infantum* and *L. braziliensis*, were also described (Peacock *et al.*, 2007). *Leishmania infantum*, also known as *L. chagasi* in Latin America, was chosen as the second *Leishmania* species to have its genome sequenced on the basis of its virulence in animals, transmissibility in sandflies and adaptability to laboratory experimentation (Denise *et al.*, 2006). This species is the causative agent of visceral leishmaniasis, the most serious form of the disease and frequently fatal if left untreated. The New World species *L. braziliensis*, within the subgenus *L. (Viannia)*, is the third and most divergent species sequenced. The *L. infantum* and *L. braziliensis* genome sequences were obtained by the whole-genome shotgun approach with five- and six-fold coverage, respectively (Peacock *et al.*, 2007). Importantly, the three complete genomes are from strains that cause distinct types of leishmanial diseases, are adapted for maintenance and manipulation in the laboratory and are also frequently used in studies *in vitro* and in animal models of infection (Laurentino *et al.*, 2004; Ivens *et al.*, 2005; Denise *et al.*, 2006). The complete sequences of all three *Leishmania* genomes can be accessed in the Tri-Tryp database; sequencing of the genome from a fourth species (*L. mexicana*) is in progress.

Surprisingly, comparative genomic studies of the three evolutionarily and geographically distinct species, *L. major* and *L. infantum* (Old World species) and *L. braziliensis* (New World species), showed very little divergence among the species in terms of genomic sequence and organization (Peacock *et al.*, 2007; Lynn and McMaster, 2008), despite a divergence of 20-100 million years within the *Leishmania* genus (Lukes *et al.*, 1997). The haploid genome of the three species has an estimated 8,300 genes, with more than 99% of them maintaining synteny in the three *Leishmania* genomes (Peacock *et al.*, 2007). As with the other Tri-Tryps, the majority of *Leishmania* genes are annotated as genes of unknown function. Of the 8,300 genes, only 200 were identified as being differentially distributed between the three genomes. The *L. braziliensis* genome possesses 47 genes that are absent from the other two species, while 27 and 5 genes are specific for the *L. infantum* and *L. major* genomes, respectively. Such high conservation in overall genome sequences and genome synteny indicates that the *Leishmania* genome is highly stable and has not undergone large genomic re-arrangements during speciation. This finding also suggests that only a few species-specific parasite genes may contribute to differential pathogenesis and tissue tropism (Peacock *et al.*, 2007; Smith *et al.*, 2007).

Thus far, the only gene for which there is experimental evidence indicating direct involvement in the differential tropism among *Leishmania* diseases is the A2 locus. Initially identified as an amastigote-specific gene family in *L. donovani*, A2 has been shown to play a major role in parasite virulence and visceralization (Zhang *et al.*, 2003). Multiple copies of the A2 gene alternating with a distinct gene termed the A2rel gene are found in the genome of several species of the *L. donovani* group that is responsible for visceral diseases. Although its precise function is still unknown, the product of the A2 gene, an endoplasmic reticulum protein with a large repetitive domain, may be related to the parasite stress response (McCall and Matlashewski, 2010). In *L. major*, which does not cause visceral leishmaniasis, the A2 gene is a pseudogene and introduction of the *L. donovani* A2 gene into *L. major* enhanced the ability of *L. major* to survive in visceral organs of susceptible BALB/c mice (Zhang *et al.*, 2003). More recently, the expression of A2 in *L. tarentolae*, a lizard parasite that is not pathogenic in mammals, significantly increased the infectivity of this species and enhanced its ability to survive in the liver of BALB/c mice (Mizbani *et al.*, 2011). In addition to experiments suggesting a possible role for the A2 gene in the differential tropism of cutaneous and visceral leishmania parasites, the A2 gene has been used as a promising vaccine candidate and diagnostic antigen (Fernandes *et al.*, 2008).

Another locus possibly involved in macrophage invasion and that varies considerably among the three *Leishmania* genomes is the GP63 locus. Also known as the major surface protease (MSP), GP63 constitutes a family of surface metalloproteases expressed in all trypanosomatids examined so far (Yao *et al.*, 2003). GP63 sequences identified in the three *Leishmania* genomes and in the *T. cruzi* and *T. brucei* genomes vary in their gene copy number, and this may have implications for differences in the disease phenotype (Voth *et al.*, 1998). A similar conclusion may apply to the amastin multi-gene family that encodes a family of amastigote-specific, highly glycosylated hydrophobic surface proteins also present in *T. cruzi* but which has been greatly expanded in the genus *Leishmania* (Teixeira *et al.*, 1995; Jackson, 2010). Interestingly, in the *T. brucei* genome, which does not have an intracellular stage, only two copies of a highly divergent amastin sequence are present (Jackson, 2010). The identification of these species-specific genes represents an initial step towards the characterization of parasite factors that may determine the specificities of each type of parasitic infection. On the other hand, antigens common to all *Leishmania* species could be used as potential vaccine candidates (Peacock *et al.*, 2007). However, apart from differences in gene sequences, it is possible that the distinct clinical manifestations observed in the different parasitic diseases may be a consequence of differential gene expression that occurs throughout the various stages of life cycle in each parasite species.

As discussed above, *Leishmania* genes are arranged in the genome as directional gene clusters that resemble prokaryotic polycistronic transcription units (Martínez-Calvillo *et al.*, 2004). This type of gene organization and polycistronic transcription have profound implications on the regulation of gene expression, which must rely on post-transcriptional mechanisms (Boucher *et al.*, 2002; Myung *et al.*, 2002; Holzer *et al.*, 2006; Leifso *et al.*, 2007). Since the dependency on promoter-based transcription initiation mechanisms for the control of mRNA levels is greatly reduced, greater emphasis is placed on post-transcriptional regulatory mechanisms controlling mRNA stability and translation, as well as protein turnover (Clayton, 2002). Despite significant differences in life stage morphology, biochemical properties and disease phenotypes, comparative gene expression studies have revealed surprisingly few differences in gene expression when mRNA levels in different life-cycle stages or in the same stage but in different *Leishmania* species were compared (Peacock *et al.*, 2007). Global interspecies analyses in *L. major* and *L. infantum* have shown that only 10%-12% of differentially expressed genes are unique to each species (Rochette *et al.*, 2008; Depledge *et al.*, 2009). A more careful examination of the protein expression patterns and the elucidation of regulatory mechanisms will provide unique insights into this important aspect of the parasite's biology. This information will also improve our understanding of the host-parasite interaction and lead to the development of new strategies for controlling leishmaniasis.

Eukaryotic genomes contain an abundance of repeated DNA and some of these repeated sequences are mobile elements. Transposable elements (TEs) are defined as DNA sequences that are able to move from one location to another in the genome and have been identified in all organisms (prokaryotic and eukaryotic) examined so far. TEs can occupy a high proportion of a species' genome. Retroposons, also known as non-long-terminal-repeat (LTR) retrotransposons, are ubiquitous elements that transpose through an RNA intermediate and are found in the genomes of most eukaryotes (Ivens *et al.*, 2005; Peacock *et al.*, 2007). *Trypanosoma brucei* and *T. cruzi* contain long autonomous retroposons of the ingi clade (Tbingi and L1Tc, respectively) and short nonautonomous truncated versions (TbRIME and NARTc, respectively), as well as degenerate ingi-related retroposons devoid of coding capacity (DIRES) that represent the most abundant transposable elements in these genomes (< 3% of the nuclear genome).

In contrast, *L. major* contains only remnants of extinct retroposons (LmDIRES) and short nonautonomous heterogenous elements (LmSIDERS). Recently, small degenerate retroposons (< 0.55 kb) containing the "79-bp signature" known as LmSIDERS (for short interspersed degenerate retroposons) have also been identified in the genomes of *L. major* (Ivens *et al.*, 2005), *L. infantum* and *L. braziliensis* (Peacock *et al.*, 2007). Unexpectedly, in *L.*

braziliensis, the site-specific non-LTR retroposon SLACS/CZAR, which is associated with tandemly repeated spliced leader sequences in an arrangement similar to that of the SLACS or CZAR element in *T. brucei* or *T. cruzi*, respectively, is fully active (Aksoy *et al.*, 1987; Villanueva *et al.*, 1991). However, in contrast to the African trypanosome genomes and similar to *L. major* (Ivens *et al.*, 2005) and *L. infantum*, no potentially active ingi-related retroposons were detected in the *L. braziliensis* genome (Bringaude *et al.*, 2009).

Mobile elements are involved in creating mutations and genomic rearrangements and, in many eukaryotes, these effects can be regulated through an RNA silencing mechanism such as RNA interference (RNAi) (Shi *et al.*, 2004b; Girard and Hannon, 2008). Since first reported in 1998 (Fire *et al.*, 1998), RNAi has swept through all fields of eukaryotic biology and has proven to be a very useful tool for analyzing gene function in a variety of organisms in which the introduction or expression of short double-stranded RNAs leads to the rapid destruction of cognate

mRNAs (Figure 4A) (Ngo *et al.*, 1998). This approach has a number of advantages: it is fast, requires very little sequence information, and reduces the expression of multiple gene copies, an action that is especially advantageous in asexual diploid organisms (LaCount *et al.*, 2000; Shi *et al.*, 2000; Wang *et al.*, 2000).

Soon after it was described in *C. elegans*, RNAi was rapidly identified in *T. brucei* (Ngo *et al.*, 1998). As discussed below, RNAi has proven to be a powerful new tool for functional genomic studies in *T. brucei*. Unexpectedly, experimental evidence as well as searches of genome databases quickly showed that RNAi is absent in *L. major* and *T. cruzi* (Robinson and Beverley, 2003; DaRocha *et al.*, 2004). It came as a surprise when Peacock *et al.* (2007) revealed that the genome of *L. braziliensis* retained key genes involved with RNAi. These authors showed that *L. braziliensis* genome contains orthologs carrying domains characteristic of the Dicer protein as well as genes with the typical argonaute domains PAZ and PIWI, the latter containing conserved amino acid residues that are essential for

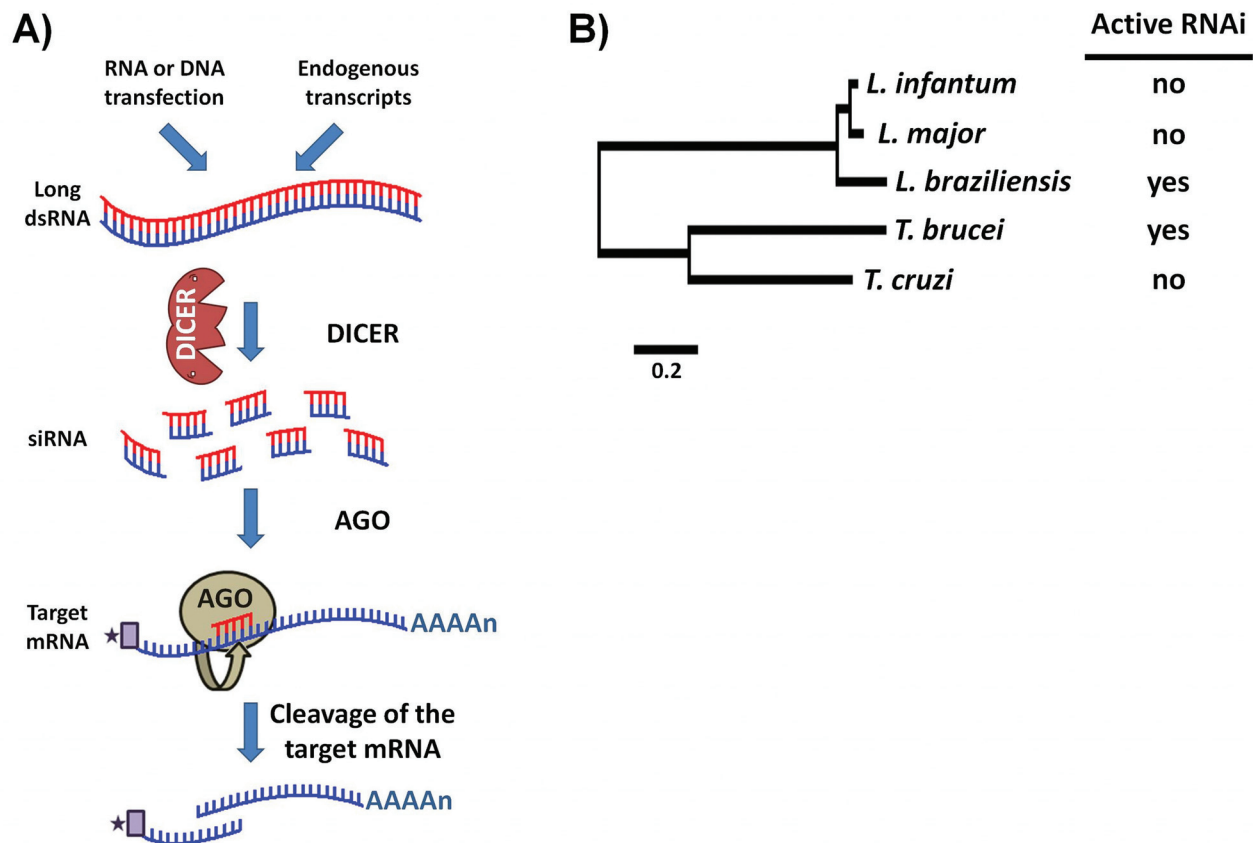


Figure 4 - Retention and loss of RNAi genes in trypanosomatids. (A) Schematic representation of the mechanism of RNA interference (RNAi) present in some trypanosomes. Double-stranded RNA (dsRNA) derived endogenously or by transfection procedures is processed by an RNase III enzyme known as DICER into small interfering dsRNA molecules (siRNA) 19-23 nt long. siRNAs associate with Argonaute (AGO), the catalytic core of the RISC (RNA-induced silencing complex), with one siRNA strand being released and the guide siRNA strand then mediating the degradation or inhibiting the translation of the target mRNA. Genes encoding components of the RNAi machinery occur only in *T. brucei* and in a sub-group of *Leishmania*. (B) Phylogenetic analysis of trypanosomatid species based on the predicted protein sequences of the housekeeping gene GAPDH (glyceraldehyde-3-phosphate dehydrogenase) and ubiquitin. The neighbor-joining tree was generated using sequences found in the parasite genome databases (www.genedb.org) and MEGA4 software. The absence or presence of genes encoding components of active RNAi machinery (as identified by Lye *et al.*, 2010) is indicated on the right.

functional TbAGO1 (Shi *et al.*, 2004a). In addition, an N-terminal RGG domain, present in TbAGO1 and shown to be essential for its association with polyribosomes, is also present in the *L. braziliensis* Ago1 gene (Shi *et al.*, 2004a). More recently, Lye *et al.* (2010) demonstrated that the RNAi pathway is functional in *L. braziliensis* and in other species within the *Leishmania* subgenus *Viannia* (*L. guyanensis* and *L. panamensis*) (Figure 4B). Thus, this divergent species of *Leishmania* appear to have retained not only the mechanisms for (RNAi)-mediated regulation but also potentially active retroposons (Peacock *et al.*, 2007), which might have assisted to create the greater divergence within the *L. braziliensis* genome compared with the other *Leishmania* species. For molecular parasitologists, these findings came as very good news since, as shown in *T. brucei*, the efficacy of RNAi knockdown as a tool for systematic analysis of gene function is now also applicable in some species of *Leishmania*.

Functional Genomics of *T. Brucei* and Gene Expression Studies in a High Throughput Era

Infection by *T. brucei* occurs after metacyclic trypomastigotes are injected into the bloodstream by the bite of a tsetse fly. In the mammalian host, the parasite differentiates into long slender trypomastigotes (bloodstream form - BSF) that divide, colonize the body fluids and can transform into a nonproliferative bloodstream form, also known as the short stump form. After a blood meal, trypomastigotes acquired by the insect vector differentiate into procyclic trypomastigotes in the gut, replicate and then migrate to salivary glands where they transform into metacyclic trypomastigotes (Matthews, 2005) (Figure 1). The ability of the BSF to invade the central nervous system leads to the neurological manifestations associated with sleeping sickness.

In contrast to *T. cruzi* and *Leishmania*, *T. brucei* develops extracellularly throughout its entire life cycle. Its direct exposure to a strong antibody response in the bloodstream requires a sophisticated immune evasion protocol, known as variant surface protein (VSG) switching (Pays *et al.*, 2007). VSGs are bloodstream-specific surface proteins with a hypervariable N-terminal domain that is exposed extracellularly and a more conserved C-terminal domain buried in the parasite's surface coat (Van der Ploeg *et al.*, 1982; Turner and Barry, 1989). Since they are tightly packed at the surface, VSGs are able to shield other invariant surface proteins from attack by the immune system. The tactic for evasion is based on antigenic variation in which the expression of a specific VSG is replaced by another gene from a supply of almost thousand copies at a rate of approximately one event per 100 cell divisions (Van der Ploeg *et al.*, 1982; Turner and Barry, 1989). VSG genes are thus a family of *T. brucei*-specific genes whose monoallelic expression distinguishes African trypanosomes from the

other two groups of pathogenic trypanosomatids; the lack of antigenic variation in *T. cruzi* and *Leishmania* spp. means that these species must evade the host's immune system by entering and multiplying inside host cells.

After completion of the *T. b. brucei* genome, it was found that of the 9,068 predicted genes, including 904 pseudogenes, there were 1,700 *T. brucei*-specific genes, 806 of them (or 9% of the whole genome) corresponding to VSGs (El-Sayed *et al.*, 2005b). Several VSG genes occur within hundreds of mini-chromosomes (50-150 kb in size) that are also part of the *T. brucei* genome. Surprisingly, analyses of all VSG sequences showed that only 57 are fully functional genes and have all the recognizable features of a typical VSG (Berriman *et al.*, 2005). It has been long known that VSG variability can be generated in *T. brucei* by creating mosaic genes that include pseudogenes (Roth *et al.*, 1989). Besides producing variability in VSG epitopes, these rearrangements may have the capacity to generate protein with new functions, as in the case of the SRA gene discussed below (Vanhamme *et al.*, 2003).

As part of their "life in the bloodstream", African trypanosomes that are pathogenic to humans have developed a mechanism to withstand other types of attack from the host immune system. In contrast to *T. brucei brucei*, which do not infect humans, *T. b. gambiense* and *T. b. rhodesiense* are resistant to trypanolytic factors (TLF) present in normal human serum (NHS). TLF activity has been attributed to apolipoprotein L1 (ApoL1) and haptoglobin (Hp)-related protein found in high density lipoprotein (for a review, see Wheeler, 2010). NHS resistance in *T. b. rhodesiense* is conferred by a truncated VSG known as serum resistance associated (SRA) protein which is located in endosomes and binds and neutralizes TLF (De Greef and Hamers, 1994; De Greef *et al.*, 1992; Vanhamme *et al.*, 2003; Pérez-Morga *et al.*, 2005; Wheeler, 2010). *Trypanosoma b. gambiense* lacks the SRA gene but still infects humans. The killing of *T. b. brucei* requires the binding of TLF-1 and trafficking to the parasite acidic lysosome. It has been proposed that, in *T. b. gambiense*, changes in the receptor that binds TLF, as well as decreased expression of this gene, are responsible for the resistance of this *T. brucei* sub-species to attack by the human innate immune system (Kieft *et al.*, 2010).

Since *T. b. brucei* is harmless to humans and *T. b. gambiense* is the most clinically relevant subspecies, Jackson *et al.* (2010) sequenced the genome of *T. b. gambiense* using the whole-genome shotgun approach combined with bacterial artificial chromosome large insert sequencing. Comparison of the sequences from a draft genome assembly of 281 contigs (~22.1 Mb) with the *T. b. brucei* genome data (Berriman *et al.*, 2005) revealed a very similar composition in terms of gene content, gene synteny and sequence identity (86.4% of *T. b. gambiense* coding sequences vary by < 1% from their orthologs in *T. b. brucei*). Even though these subspecies behave differently in humans, there were

very few differences in their genomes. One of these differences involved a gene encoding a putative iron-ascorbate oxidoreductase that is specific to a few *T. b. brucei* strains and is also absent in other trypanosomatids (*L. major* and *T. cruzi*). Thus, it seems that not only differences in gene content *per se*, but also individual single nucleotide polymorphisms (indels) and variations in gene expression, or a combination of these factors, may contribute to this phenotypic variation (Jackson *et al.*, 2010).

Coordinated changes in gene expression are vital for trypanosomes since they must deal with rapid changes in their environment (nutrient availability, temperature, host defenses, the presence of drug, etc.) and be able to disseminate by alternating through different hosts. In addition, unique mechanisms for regulating gene expression are required since there is an almost complete absence of transcriptional control at the level of initiation. The lack of transcriptional regulatory elements, including a typical RNA polymerase II promoter, led to the conclusion that most factors involved in controlling gene expression act at the post-transcriptional level. By using approaches such as differential display, RNA fingerprinting, differential screening of cDNA libraries, random sequencing of cDNA clones and DNA microarrays, several groups have shown that significant changes in mRNA levels occur during the life cycle of all Tri-Tryps (Teixeira *et al.*, 1995; El-Sayed *et al.*, 1995; Mathieu-Daudé *et al.*, 1996; Diehl *et al.*, 2002; Cerqueira *et al.*, 2005; Kabani *et al.*, 2009). Experimental evidence also indicates that, in addition to mRNA stability, changes in polysomal mobilization constitute an important mechanism for regulating gene expression (Alves *et al.*, 2010). With the advent of next generation sequencing technologies such as RNAseq, a global description of gene expression patterns in *T. brucei* has finally been achieved (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010; Veitch *et al.*, 2010). This approach has provided much more complete information about mRNA structure and expression levels, including the patterns of spliced leader (SL) and poly-A additions, as well as alternative pre-mRNA processing. Moreover, these global gene expression analyses have confirmed that only two *T. brucei* genes (poly-A polymerase and DNA/RNA helicase) contain introns.

Nilsson *et al.* (2010) showed that 2,500 alternative splicing events occur during processing of *T. brucei* genes, with a large number of these being regulated during the life cycle (a total of 600 genes have transcripts with more than one trans-spliced variant). The alternative splicing reactions can alter the message dramatically, as shown by Nilsson *et al.* (2010) and represented in Figure 3, with changes in the SL addition site leading to alterations in regulatory elements in the 5'UTR (resulting in the modification of gene expression) or the initiator AUG (generating a different N-terminus or even causing a complete change in the translated ORF). Aminoacyl tRNA synthetase tran-

scripts are interesting examples of trans-spliced variants since differences in the enzyme N-terminus result in changes in the protein targeting signal: distinct mRNAs produce proteins that are directed to mitochondria or remain in the cytoplasm (Nilsson *et al.*, 2010). Likewise, alternative splicing is a potential mechanism for dual localization of the *T. cruzi* LYT1 gene (Benabdellah *et al.*, 2007). In addition to providing a mechanism for generating changes in the proteome, such flexibility in the selection of splicing sites is compatible with evidence that polycistronic transcription may also initiate at internal sites in gene clusters since correct processing of the 5' end of mRNA can occur at various points in the primary transcript (Kolev *et al.*, 2010).

The mechanisms involved in transcription initiation in trypanosomes have always been an intriguing question. Recently, chromatin immunoprecipitation in combination with conventional Sanger sequencing (Respuela *et al.*, 2008) or next-generation sequencing technology (Siegel *et al.*, 2009; Thomas *et al.*, 2009; Wright *et al.*, 2010) has provided a much awaited analysis of the distribution patterns of modified histones throughout the *T. brucei* genome. As summarized in Figure 3, the divergent strand-switch regions (SSR) have been found to contain an enrichment of histone variants (H2AZ and H2BV), acetylated histone 4 at lysine 10 (H4K10ac), histone 3 acetylated at residues K9 and K14 and trimethylated at K4, the bromodomain factor BDF3, and the transcription factors TRF4 and SNAP50 (Respuela *et al.*, 2008; Siegel *et al.*, 2009; Thomas *et al.*, 2009; Wright *et al.*, 2010). The histone variants H3V and H4V are present at transcription termination sites (Siegel *et al.*, 2009). Similar patterns of chromatin modifications from transcription start sites (TSS) located at the beginning of polycistronic units were also found internally in the same polycistronic unit, suggesting the presence of internal TSS, in agreement with RNAseq data (Siegel *et al.*, 2009; Kolev *et al.*, 2010; Wright *et al.*, 2010).

The increasing number of published genomes and transcriptomic data, partly as a consequence of the introduction of new sequencing platforms, has created enormous challenges for the field of functional genomics. One of the most useful techniques that has been used to identify gene function in *T. brucei* is a combination of the inducible system mediated by T7 RNA polymerase and the tetracycline repressor (Wirtz *et al.*, 1999) in combination with RNAi gene knockdown. As indicated before, in contrast to *T. cruzi* and *L. major*, RNAi is functional in *T. brucei* (Ngo *et al.*, 1998; Robinson and Beverley 2003; DaRocha *et al.*, 2004; Lye *et al.*, 2010). The first large scale functional genomics study using RNAi was reported by Morris *et al.* (2002), who transfected *T. brucei* procyclic forms with a random RNAi library. This library was generated by using a construct to inducibly express dsRNA via two head-to-head tetracycline-regulated T7 RNA polymerase promoters. By selecting parasites that expressed the dsRNA

but were unable to bind lectin, these authors identified clones with a reduced glycosylation of surface proteins and showed that these clones had lower expression of hexokinase (Morris *et al.*, 2002).

A few years later, the functions of 197 ORFs from *T. b. brucei* chromosome I were tested by RNAi. RNAi-induced parasites were tested for growth, nuclear and kinetoplast abnormalities by DAPI staining, and a pleiotropic phenotype, morphology and motility. At least one of these phenotypes was found in 68 individual knockdowns (Subramaniam *et al.*, 2006). In a similar approach, Portman *et al.* (2009) identified novel components of the paraflagellar rod (PFR) structure by comparing proteomic profiles before and after ablating the expression of individual PFR proteins. The silencing of PFR1 and PFR15 was evaluated by two-dimensional difference gel electrophoresis, and the spots with a two-fold change in volume were subjected to tandem MS protein identification. Proteomic analysis combined with RNAi knockdown led to the identification of 30 proteins as potential PFR components, 20 of which were novel proteins. High-throughput cloning systems also enhance functional analyses. By using the Gateway® technology to create yeast two-hybrid vectors, eight non-redundant protein-protein interactions were detected among proteins with a PFR structure (Lacomble *et al.*, 2009). These authors were able to construct a map showing the complex interaction of PFR proteins.

The availability of efficient methods for genetic manipulation and for testing gene function through RNAi knockdown has led to major advances in genomic, transcriptomic and proteomic analyses of *T. brucei*, and has made this species a model organism for studying basic aspects of trypanosomatid biology. More recently, with the discovery of functional RNAi machinery in *L. braziliensis* and other *Leishmania* species, gene function studies can now be conducted in this group and will soon provide valuable new information about *Leishmania*-specific genes. However, as is becoming increasingly apparent from the data generated by comparative genomic analyses, each of the trypanosomatid species has its peculiarities. Researchers thus face the challenge of developing new protocols specific for studying each parasite and its corresponding diseases. With our current knowledge of each Tri-Tryp disease and the new research methods that are being developed we can expect many new studies to emerge from hypotheses based on the Tri-Tryp genomic data. As a consequence of these new findings, better methods of controlling and preventing these diseases will follow.

Acknowledgments

The work from SMT and RMCP was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), Instituto Nacional de Ciência e Tecnologia de Vacinas (INCTV) and the Howard

Hughes Medical Institute (HHMI). The work from WDR and MMKM was supported by FAPEMIG, Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Paraná (Fundação Araucária), CAPES/Reuni, PPSUS/MS and CNPq.

References

- Aksoy S, Lalor TM, Martin J, Van der Ploeg LH and Richards FF (1987) Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. EMBO J 6:3819-3826.
- Alves LR, Avila AR, Correa A, Holetz FB, Mansur FC, Manque PA, de Menezes JP, Buck GA, Krieger MA and Goldenberg S (2010) Proteomic analysis reveals the dynamic association of proteins with translated mRNAs in *Trypanosoma cruzi*. Gene 452:72-78.
- Araujo PR, Burle-Caldas GA, Silva-Pereira RA, Bartholomeu DC, DaRocha WD and Teixeira SM (2011) Development of a dual reporter system to identify regulatory cis-acting elements in untranslated regions of *Trypanosoma cruzi* mRNAs. Parasitol Int 60:161-169.
- Arner E, Kindlund E, Nilsson D, Farzana F, Ferella M, Tammi MT and Andersson B (2007) Database of *Trypanosoma cruzi* repeated genes: 20,000 additional gene variants. BMC Genomics 8:e391.
- Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, *et al.* (2010) TriTrypDB: A functional genomic resource for the Trypanosomatidae. Nucleic Acids Res 38:D457-D462.
- Augusto-Pinto L, Teixeira SM, Pena SD and Machado CR (2003) Single-nucleotide polymorphisms of the *Trypanosoma cruzi* MSH2 gene support the existence of three phylogenetic lineages presenting differences in mismatch-repair efficiency. Genetics 164:117-126.
- Baida RC, Santos MR, Carmo MS, Yoshida N, Ferreira D, Ferreira AT, El Sayed NM, Andersson B and da Silveira JF (2006) Molecular characterization of serine-, alanine-, and proline-rich proteins of *Trypanosoma cruzi* and their possible role in host cell infection. Infect Immun 74:1537-1546.
- Bartholomeu DC, Cerqueira GC, Leão AC, DaRocha WD, Pais FS, Macedo C, Djikeng A, Teixeira SM and El-Sayed NM (2009) Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. Nucleic Acids Res 37:3407-3417.
- Bates PA (2007) Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. Int J Parasitol 37:1097-1106.
- Benabdellah K, González-Rey E and González A (2007) Alternative trans-splicing of the *Trypanosoma cruzi* LYT1 gene transcript results in compartmental and functional switch for the encoded protein. Mol Microbiol 65:1559-1567.
- Benz C, Nilsson D, Andersson B, Clayton C and Guilbride DL (2005) Messenger RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol 143:125-134.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. Science 309:416-422.

- Blackwell JM and Melville SE (1999) Status of protozoan genome analysis: Trypanosomatids. *Parasitology* 118 (Suppl):S11-14.
- Blackwell JM, Fakiola M, Ibrahim ME, Jamieson SE, Jeronimo SB, Miller EN, Mishra A, Mohamed HS, Peacock CS, Raju M, *et al.* (2009) Genetics and visceral leishmaniasis: Of mice and man. *Parasite Immunol* 31:254-266.
- Boucher N, McNicoll F, Dumas C and Papadopoulos B (2002) RNA polymerase I-mediated transcription of a reporter gene integrated into different loci of *Leishmania*. *Mol Biochem Parasitol* 119:153-158.
- Branche C, Ochaya S, Aslund L and Andersson B (2006) Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 147:30-38.
- Brandão A, Urmenyi T, Rondinelli E, Gonzalez A, de Miranda AB and Degraive W (1997) Identification of transcribed sequences (ESTs) in the *Trypanosoma cruzi* genome project. *Mem Inst Oswaldo Cruz* 92:863-866.
- Brener Z (1973) Biology of *Trypanosoma cruzi*. *Annu Rev Microbiol* 27:347-382.
- Bringaud F, Berriman M and Hertz-Fowler C (2009) Trypanosomatid genomes contain several subfamilies of ingi-related retrotransposons. *Eukaryot Cell* 8:1532-1542.
- Brisse S, Dujardin JC and Tibayrenc M (2000) Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. *Mol Biochem Parasitol* 111:95-105.
- Buscaglia CA and Di Noia JM (2003) *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. *Microbes Infect* 5:419-427.
- Campos PC, Bartholomeu DC, DaRocha WD, Cerqueira GC and Teixeira SM (2008) Sequences involved in mRNA processing in *Trypanosoma cruzi*. *Int J Parasitol* 38:1383-1389.
- Cano MI, Gruber A, Vazquez M, Cortés A, Levin MJ, González A, Degraive W, Rondinelli E, Zingales B, Ramirez JL, *et al.* (1995) Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project. *Mol Biochem Parasitol* 71:273-278.
- Cerqueira GC, DaRocha WD, Campos PC, Zouain CS and Teixeira SM (2005) Analysis of expressed sequence tags from *Trypanosoma cruzi* amastigotes. *Mem Inst Oswaldo Cruz* 100:385-389.
- Cerqueira GC, Bartholomeu DC, DaRocha WD, Hou L, Freitas-Silva DM, Machado CR, El-Sayed NM and Teixeira SM (2008) Sequence diversity and evolution of multigene families in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 157:65-72.
- Clayton CE (2002) Life without transcriptional control? From fly to man and back again. *EMBO J* 21:1881-1888.
- Cribb P and Serra E (2009) One- and two-hybrid analysis of the interactions between components of the *Trypanosoma cruzi* spliced leader RNA gene promoter binding complex. *Int J Parasitol* 39:525-532.
- Cribb P, Esteban L, Trochine A, Girardini J and Serra E (2010) *Trypanosoma cruzi* TBP shows preference for C/G-rich DNA sequences *in vitro*. *Exp Parasitol* 124:346-349.
- DaRocha WD, Otsu K, Teixeira SM and Donelson JE (2004) Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 133:175-186.
- de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, Teixeira SM, Chiari E, Junqueira AC, Fernandes O, Macedo AM, *et al.* (2006) Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathog* 2:e24.
- De Greef C and Hamers R (1994) The serum resistance-associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. *Mol Biochem Parasitol* 68:277-284.
- De Greef C, Chimfwembe E, Kihang'a Wabacha J, Bajanya Songa E and Hamers R (1992) Only the serum-resistant bloodstream forms of *Trypanosoma brucei rhodesiense* express the serum resistance associated (SRA) protein. *Ann Soc Belg Med Trop* 72(Suppl 1):13-21.
- Denise H, Poot J, Jimenez M, Ambit A, Hermman DC, Vermeulen AN, Coombs GH and Mottram JC (2006) Studies on the CPA cysteine peptidase in the *Leishmania infantum* genome strain JPCM5. *BMC Mol Biol* 7:42.
- Depledge DP, Evans KJ, Ivens AC, Aziz N, Maroof A, Kaye PM and Smith DF (2009) Comparative expression profiling of *Leishmania*: Modulation in gene expression between species and in different host genetic backgrounds. *PLoS Negl Trop Dis* 3:e476.
- Desjeux P (1996) Leishmaniasis. Public health aspects and control. *Clin Dermatol* 14:417-423.
- Di Noia JM, Sánchez DO and Frasch AC (1995) The protozoan *Trypanosoma cruzi* has a family of genes resembling the mucin genes of mammalian cells. *J Biol Chem* 270:24146-24149.
- Diehl S, Diehl F, El-Sayed NM, Clayton C and Hoheisel JD (2002) Analysis of stage-specific gene expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA-microarray. *Mol Biochem Parasitol* 123:115-123.
- El-Sayed NM and Donelson JE (1997) A survey of the *Trypanosoma brucei rhodesiense* genome using shotgun sequencing. *Mol Biochem Parasitol* 84:167-178.
- El-Sayed NM, Alarcon CM, Beck JC, Sheffield VC and Donelson JE (1995) cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. *Mol Biochem Parasitol* 73:75-90.
- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, *et al.* (2005a) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309:404-409.
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, *et al.* (2005b) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309:409-415.
- Fernandes AP, Costa MM, Coelho EA, Michalick MS, de Freitas E, Melo MN, Luiz Tafuri W, Resende DM, Hermont V, Abrantes CF, *et al.* (2008) Protective immunity against challenge with *Leishmania (Leishmania) chagasi* in beagle dogs vaccinated with recombinant A2 protein. *Vaccine* 26:5888-5895.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.
- Franzén O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA and Andersson B (2011) Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. *PLoS Negl Trop Dis* 5:e984.

- Girard A and Hannon GJ (2008) Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol* 18:136-148.
- Gunzl A, Bruderer T, Laufer G, Schimanski B, Tu LC, Chung HM, Lee PT and Lee MG (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell* 2:542-551.
- Haag J, O'hUigin C and Overath P (1998) The molecular phylogeny of trypanosomes: Evidence for an early divergence of the Salivaria. *Mol Biochem Parasitol* 91:37-49.
- Hajduk SL, Harris ME and Pollard VW (1993) RNA editing in kinetoplastid mitochondria. *FASEB J* 7:54-63.
- Henriksson J, Porcel B, Rydåker M, Ruiz A, Sabaj V, Galanti N, Cazzulo JJ, Frasch AC and Pettersson U (1995) Chromosome specific markers reveal conserved linkage groups in spite of extensive chromosomal size variation in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 73:63-74.
- Holzer TR, McMaster WR and Forney JD (2006) Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in *Leishmania mexicana*. *Mol Biochem Parasitol* 146:198-218.
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309:436-442.
- Jackson AP (2010) The evolution of amastin surface glycoproteins in trypanosomatid parasites. *Mol Biol Evol* 27:33-45.
- Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, Chukualim B, Capewell P, MacLeod A, Melville SE, *et al.* (2010) The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human African trypanosomiasis. *PLoS Negl Trop Dis* 4:e658.
- Kabani S, Fenn K, Ross A, Ivens A, Smith TK, Ghazal P and Matthews K (2009) Genome-wide expression profiling of *in vivo*-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*. *BMC Genomics* 10:e427.
- Kieft R, Capewell P, Turner CM, Veitch NJ, MacLeod A and Hajduk S (2010) Mechanism of *Trypanosoma brucei gambiense* (group 1) resistance to human trypanosome lytic factor. *Proc Natl Acad Sci USA* 107:16137-16141.
- Koeller CM and Heise N (2011) The sphingolipid biosynthetic pathway is a potential target for chemotherapy against Chagas disease. *Enzyme Res* 2011:e648159.
- Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S and Tschudi C (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog* 6:e1001090.
- Lacomble S, Portman N and Gull K (2009) A protein-protein interaction map of the *Trypanosoma brucei* paraflagellar rod. *PLoS One* 4:e7685.
- LaCount DJ, Bruse S, Hill KL and Donelson JE (2000) Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. *Mol Biochem Parasitol* 111:67-76.
- Lainson R, Ward RD and Shaw JJ (1977) *Leishmania* in phlebotomid sandflies: VI. Importance of hindgut development in distinguishing between parasites of the *Leishmania mexicana* and *L. braziliensis* complexes. *Proc R Soc Lond B* 199:309-320.
- Laurentino EC, Ruiz JC, Fazelini G, Myler PJ, Degraive W, Ferreira MA, Ribeiro LMC and Cruz AK (2004) A survey of *Leishmania braziliensis* genome by shotgun sequencing. *Mol Biochem Parasitol* 13:81-86.
- LeBowitz JH, Smith HQ, Rusche L and Beverley SM (1993) Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* 7:996-1007.
- Leifso K, Cohen-Freue G, Dogra N, Murray A and McMaster WR (2007) Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: The *Leishmania* genome is constitutively expressed. *Mol Biochem Parasitol* 152:35-46.
- Lepesheva GI, Villalta F and Waterman MR (2011) Targeting *Trypanosoma cruzi* sterol 14 α -demethylase (CYP51). *Adv Parasitol* 75:65-87.
- Levick MP, Blackwell JM, Connor V, Coulson RM, Miles A, Smith HE, Wan KL and Ajioka JW (1996) An expressed sequence tag analysis of a full-length, spliced-leader cDNA library from *Leishmania major* promastigotes. *Mol Biochem Parasitol* 76:345-348.
- Liang XH, Haritan A, Uliel S and Michaeli S (2003) Trans and cis splicing in trypanosomatids: Mechanism, factors, and regulation. *Eukaryot Cell* 2:830-840.
- Lima MT, Lenzi HL and Gattass CR (1995) Negative tissue parasitism in mice injected with a noninfective clone of *Trypanosoma cruzi*. *Parasitol Res* 81:6-12.
- Lukes J, Jirků M, Dolezel D, Kral'ová I, Hollar L and Maslov DA (1997) Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. *J Mol Evol* 44:521-527.
- Lye LF, Owens K, Shi H, Murta SM, Vieira AC, Turco SJ, Tschudi C, Ullu E and Beverley SM (2010) Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6:e1001161.
- Lynn MA and McMaster RW (2008) *Leishmania*: Conserved evolution – Diverse diseases. *Trends Parasitol* 24:103-105.
- Martínez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K and Myler PJ (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* 11:1291-1299.
- Martínez-Calvillo S, Nguyen D, Stuart K and Myler PJ (2004) Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot Cell* 3:506-517.
- Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG and Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. *J Biomed Biotechnol* 2010:525241.
- Mathieu-Daudé F, Cheng R, Welsh J and McClelland M (1996) Screening of differentially amplified cDNA products from RNA arbitrarily primed PCR fingerprints using single strand conformation polymorphism (SSCP) gels. *Nucleic Acids Res* 24:1504-1507.
- Matthews KR (2005) The developmental cell biology of *Trypanosoma brucei*. *J Cell Sci* 118:283-290.
- Mauel J (2002) Vaccination against *Leishmania* infections. *Curr Drug Targets Immune Endocr Metabol Disord* 2:201-226.
- McCall LI and Matlashewski G (2010) Localization and induction of the A2 virulence factor in *Leishmania*: Evidence that A2 is a stress response protein. *Mol Microbiol* 77:518-530.
- Miles MA, Souza A, Povoas M, Shaw JJ, Lainson R and Tøye PJ (1978) Isozymic heterogeneity of *Trypanosoma cruzi* in the

- first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature* 272:819-821.
- Mizbani A, Taslimi Y, Zahedifard F, Taheri T and Rafati S (2011) Effect of A2 gene on infectivity of the nonpathogenic parasite *Leishmania tarentolae*. *Parasitol Res* 109:793-799.
- Momen H (1999) Taxonomy of *Trypanosoma cruzi*: A commentary on characterization and nomenclature. *Mem Inst Oswaldo Cruz* 94(Suppl 1):181-184.
- Morris JC, Wang Z, Drew ME and Englund PT (2002) Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. *EMBO J* 21:4429-4438.
- Murray HW, Berman JD, Davies CR and Saravia NG (2005) Advances in leishmaniasis. *Lancet* 366:1561-1577.
- Myler PJ, Audleman L, deVos T, Hixson G, Kiser P, Lemley C, Magness C, Rickel E, Sisk E, Sunkin S, et al. (1999) *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci USA* 96:2902-2906.
- Myung KS, Beetham JK, Wilson ME and Donelson JE (2002) Comparison of the post-transcriptional regulation of the mRNAs for the surface proteins PSA (GP46) and MSP (GP63) of *Leishmania chagasi*. *J Biol Chem* 277:16489-16497.
- Ngo H, Tschudi C, Gull K and Ullu E (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci USA* 95:14687-14692.
- Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I and Ochsenreiter T (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog* 6:e1001037.
- Olin-Sandoval V, Moreno-Sánchez R and Saavedra E (2010) Targeting trypanothione metabolism in trypanosomatid human parasites. *Curr Drug Targets* 11:1614-1630.
- Pays E, Salmon D, Morrison LJ, Marcello L and Barry JD (2007) Antigenic variation in *Trypanosoma brucei*. In: Barry JD, McCulloch R, Mottram J and Acosta-Serrano A (eds) *Trypanosomes After the Genome*. Horizon Bioscience, Wymondham, pp 339-372.
- Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M, et al. (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* 39:839-847.
- Pérez-Morga D, Vanhollebeke B, Paturiaux-Hanocq F, Nolan DP, Lins L, Homblé F, Vanhamme L, Tebabi P, Pays A, Poelvoorde P, et al. (2005) Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science* 309:469-472.
- Porcel BM, Tran AN, Tammi M, Nyarady Z, Rydåker M, Urmenyi TP, Rondinelli E, Pettersson U, Andersson B and Aslund L (2000) Gene survey of the pathogenic protozoan *Trypanosoma cruzi*. *Genome Res* 10:1103-1107.
- Portman N, Lacomble S, Thomas B, McKean PG and Gull K (2009) Combining RNA interference mutants and comparative proteomics to identify protein components and dependencies in a eukaryotic flagellum. *J Biol Chem* 284:5610-5619.
- Respuela P, Ferella M, Rada-Iglesias A and Aslund L (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. *J Biol Chem* 283:15884-15892.
- Robinson KA and Beverley SM (2003) Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol* 128:217-228.
- Rochette A, Raymond F, Ubéda JM, Smith M, Messier N, Boissvert S, Rigault P, Corbeil J, Ouellette M and Papadopoulos B (2008) Genome-wide gene expression profiling analysis of *Leishmania major* and *Leishmania infantum* developmental stages reveals substantial differences between the two species. *BMC Genomics* 9:255.
- Roth C, Bringaud F, Layden RE, Baltz T and Eisen H (1989) Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc Natl Acad Sci USA* 86:9375-9379.
- Ruvalcaba-Trejo LI and Sturm NR (2011) The *Trypanosoma cruzi* Sylvio X10 strain maxicircle sequence: The third musketeer. *BMC Genomics* 12:58.
- Sakthianandeswaren A, Foote SJ and Handman E (2009) The role of host genetics in leishmaniasis. *Trends Parasitol* 25:383-391.
- Shi H, Djikeng A, Tschudi C and Ullu E (2004) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: Control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol* 24:420-427.
- Shi H, Djikeng A, Mark T, Wirtz E, Tschudi C and Ullu E (2000) Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *RNA* 6:1069-1076.
- Shi H, Chamond N, Tschudi C and Ullu E (2004) Selection and characterization of RNA interference-deficient trypanosomes impaired in target mRNA degradation. *Eukaryot Cell* 3:1445-1453.
- Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S and Cross GA (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* 23:1063-1076.
- Siegel TN, Hekstra DR, Wang X, Dewell S and Cross GA (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res* 38:4946-4957.
- Simpson AG, Stevens JR and Lukes J (2006) The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol* 22:168-174.
- Smith DF, Peacock CS and Cruz AK (2007) Comparative genomics: From genotype to disease phenotype in the leishmaniasis. *Int J Parasitol* 37:1173-1186.
- Smith M, Blanchette M and Papadopoulos B (2008) Improving the prediction of mRNA extremities in the parasitic protozoan *Leishmania*. *BMC Bioinformatics* 9:158.
- Souto RP, Fernandes O, Macedo AM, Campbell DA and Zingales B (1996) DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 83:141-152.
- Stevens JR, Noyes HA, Dover GA and Gibson WC (1999) The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology* 118:107-116.
- Stuart K and Panigrahi AK (2002) RNA editing: Complexity and complications. *Mol Microbiol* 45:591-596.
- Subramaniam C, Veazey P, Redmond S, Hayes-Sinclair J, Chambers E, Carrington M, Gull K, Matthews K, Horn D and

- Field MC (2006) Chromosome-wide analysis of gene function by RNA interference in the African trypanosome. *Eukaryot Cell* 5:1539-1549.
- Teixeira SM, Kirchhoff LV and Donelson JE (1995) Post-transcriptional elements regulating expression of mRNAs from the amastin/tuzin gene cluster of *Trypanosoma cruzi*. *J Biol Chem* 270:22586-22594.
- Thomas S, Green A, Sturm NR, Campbell DA and Myler PJ (2009) Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics* 10:e152.
- Turner CM and Barry JD (1989) High frequency of antigenic variation in *Trypanosoma brucei rhodesiense* infections. *Parasitology* 99:67-75.
- Van der Ploeg LH, Valerio D, De Lange T, Bernards A, Borst P and Grosveld FG (1982) An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic Acids Res* 10:5905-5923.
- Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, Nolan DP, Lins L, Van Den Abbeele J, Pays A, Tebabi P, Van Xong H, Jacquet A, *et al.* (2003) Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* 422:83-87.
- Vargas N, Pedrosa A and Zingales B (2004) Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Mol Biochem Parasitol* 138:131-141.
- Veitch NJ, Johnson PC, Trivedi U, Terry S, Wildridge D and MacLeod A (2010) Digital gene expression analysis of two life cycle stages of the human-infective parasite, *Trypanosoma brucei gambiense*, reveals differentially expressed clusters of co-regulated genes. *BMC Genomics* 11:e124.
- Verdun RE, Di Paolo N, Urmenyi TP, Rondinelli E, Frasch AC and Sanchez DO (1998) Gene discovery through expressed sequence Tag sequencing in *Trypanosoma cruzi*. *Infect Immun* 66:5393-5398.
- Villanueva MS, Williams SP, Beard CB, Richards FF and Aksoy S (1991) A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. *Mol Cell Biol* 11:6139-6148.
- Voth BR, Kelly BL, Joshi PB, Ivens AC and McMaster WR (1998) Differentially expressed *Leishmania major* gp63 genes encode cell surface leishmanolysin with distinct signals for glycosylphosphatidylinositol attachment. *Mol Biochem Parasitol* 93:31-41.
- Wang Z, Morris JC, Drew ME and Englund PT (2000) Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *J Biol Chem* 275:40174-40179.
- Weatherly DB, Boehlke C and Tarleton RL (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics* 10:e255.
- Westenberger SJ, Cerqueira GC, El-Sayed NM, Zingales B, Campbell DA and Sturm NR (2006) *Trypanosoma cruzi* mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics* 7:e60.
- Wheeler RJ (2010) The trypanolytic factor-mechanism, impacts and applications. *Trends Parasitol* 26:457-464.
- Wincker P (1996) The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res* 24:1688-1694.
- Wirtz E, Leal S, Ochatt C and Cross GA (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol* 99:89-101.
- Worthey EA, Martinez-Calvillo S, Schnauffer A, Aggarwal G, Cawthra J, Fazelinia G, Fong C, Fu G, Hassebrock M, Hixson G, *et al.* (2003) *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Res* 31:4201-4210.
- Wright AD, Li S, Feng S, Martin DS and Lynn DH (1999) Phylogenetic position of the kinetoplasts, *Cryptobia bullocki*, *Cryptobia catostomi*, and *Cryptobia salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences. *Mol Biochem Parasitol* 99:69-76.
- Wright JR, Siegel TN and Cross GA (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Mol Biochem Parasitol* 172:141-144.
- Yao C, Donelson JE and Wilson ME (2003) The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Mol Biochem Parasitol* 132:1-16.
- Zhang WW, Mendez S, Ghosh A, Myler P, Ivens A, Clos J, Sacks DL and Matlashewski G (2003) Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. *J Biol Chem* 278:35508-35515.
- Zhou S, Kile A, Kvikstad E, Bechner M, Severin J, Forrest D, Runnheim R, Churas C, Anantharaman TS, Myler P, *et al.* (2004) Shotgun optical mapping of the entire *Leishmania major* Friedlin genome. *Mol Biochem Parasitol* 138:97-106.
- Zingales B, Pereira ME, Almeida KA, Umezawa ES, Nehme NS, Oliveira RP, Macedo A and Souto RP (1997) Biological parameters and molecular markers of clone CL Brener – The reference organism of the *Trypanosoma cruzi* genome project. *Mem Inst Oswaldo Cruz* 92:811-814.
- Zingales B, Souto RP, Mangia RH, Lisboa CV, Campbell DA, Coura JR, Jansen A and Fernandes O (1998) Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. *Int J Parasitol* 28:105-112.
- Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, *et al.* (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: Second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz* 104:1051-1054.

Associate Editor: Carlos F.M. Menck

ANEXO 2

RESEARCH ARTICLE

Open Access

Distinct genomic organization, mRNA expression and cellular localization of members of two amastin sub-families present in *Trypanosoma cruzi*

Monica Mendes Kangussu-Marcolino¹, Rita Márcia Cardoso de Paiva², Patrícia Rosa Araújo², Rondon Pessoa de Mendonça-Neto², Laiane Lemos¹, Daniella Castanheira Bartholomeu³, Renato A Mortara⁴, Wanderson Duarte da Rocha^{1*} and Santuza Maria Ribeiro Teixeira^{2*}

Abstract

Background: Amastins are surface glycoproteins (approximately 180 residues long) initially described in *Trypanosoma cruzi* as particularly abundant during the amastigote stage of this protozoan parasite. Subsequently, they have been found to be encoded by large gene families also present in the genomes of several species of *Leishmania* and in other Trypanosomatids. Although most amastin genes are organized in clusters associated with tuzin genes and are up-regulated in the intracellular stage of *T. cruzi* and *Leishmania* spp, distinct genomic organizations and mRNA expression patterns have also been reported.

Results: Based on the analysis of the complete genome sequences of two *T. cruzi* strains, we identified a total of 14 copies of amastin genes in *T. cruzi* and showed that they belong to two of the four previously described amastin subfamilies. Whereas δ -amastin genes are organized in two or more clusters with alternating copies of tuzin genes, the two copies of β -amastins are linked together in a distinct chromosome. Most *T. cruzi* amastins have similar surface localization as determined by confocal microscopy and western blot analyses. Transcript levels for δ -amastins were found to be up-regulated in amastigotes from several *T. cruzi* strains, except in the G strain, which is known to have low infection capacity. In contrast, in all strains analysed, β -amastin transcripts are more abundant in epimastigotes, the stage found in the insect vector.

Conclusions: Here we showed that not only the number and diversity of *T. cruzi* amastin genes is larger than what has been predicted, but also their mode of expression during the parasite life cycle is more complex. Although most *T. cruzi* amastins have a similar surface localization, only δ -amastin genes have their expression up-regulated in amastigotes. The results showing that a sub-group of this family is up-regulated in epimastigotes, suggest that, in addition of their role in intracellular amastigotes, *T. cruzi* amastins may also serve important functions during the insect stage of the parasite life cycle. Most importantly, evidence for their role as virulence factors was also unveiled from the data showing that δ -amastin expression is down regulated in a strain presenting low infection capacity.

Keywords: *Trypanosoma cruzi*, Amastigote, Amastin, mRNA

* Correspondence: wanderson_darocha@yahoo.com; santuzat@icb.ufmg.br

¹Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Rua Quinze de Novembro, 1299, Centro Curitiba, PR 80060-000, Brazil

²Departamento de Bioquímica e Imunologia, Av. Antônio Carlos, 6627, Pampulha Belo Horizonte, MG 31270-901, Brazil

Full list of author information is available at the end of the article

Background

Trypanosoma cruzi, the protozoan parasite that is the etiologic agent of Chagas disease [1], undergoes four developmental stages during its complex life cycle: epimastigotes and metacyclic trypomastigotes, present in the insect vector, and intracellular amastigotes and bloodstream trypomastigotes, present in the mammalian host. This parasite must rely on a broad set of genes that allow it to multiply in the insect gut, to differentiate into forms that are able to invade and multiply inside a large number of distinct mammalian cell types and to circumvent the host immune system. To meet the challenges it faces during its life cycle, complex regulatory mechanisms must control the expression of the *T. cruzi* repertoire of about 12,000 genes. Among them, there are several large gene families encoding surface proteins, which are key players directly involved in host-parasite interactions (reviewed by Epting et al. [2]).

The amastin gene family was initially reported as a group of *T. cruzi* genes encoding 174 amino acid transmembrane glycoproteins and whose mRNA are 60-fold more abundant in amastigotes than in epimastigotes or trypomastigotes [3]. The differential expression of amastin mRNAs during the *T. cruzi* life cycle has been attributed to cis-acting elements present in the 3'UTR as well as to RNA binding proteins that may recognize this sequence [4,5]. It is also known that amastin genes alternate with genes encoding a cytoplasmic protein named tuzin [6]. After the completion of the genome sequences of several Trypanosomatids it was revealed that the amastin gene family is also present in various *Leishmania* species as well as in two related insect parasites, *Leptomonas seymouri* and *Crithidia* spp [7-9]. It has also been reported that this gene family is actually much larger in the genus *Leishmania* when compared to other Trypanosomatids. Predicted topology based on sequences found in the genomes of *L. major*, *L. infantum* and *T. cruzi* indicates that all amastins have four transmembrane regions, two extracellular domains and N- and C-terminal tails facing the cytosol [8]. Moreover, comparative analyses of amastin genes belonging to six *T. cruzi* strains evidenced that sequences encoding the hydrophilic, extracellular domain, which is less conserved, have higher intragenomic variability in strains belonging to *T. cruzi* group II and hybrid strains compared to *T. cruzi* I strains [10]. Based on phylogenetic analyses of amastin orthologs from various Trypanosomatids, it has been proposed that amastins can be classified into four subfamilies, named α -, β -, γ -, and δ - amastins. Importantly, in *L. major* and *L. infantum*, in which members of all four sub-families are found, amastin genes showed differences in genomic positions and expression patterns of their mRNAs [8,9].

More than fifteen years after their discovery, the function of amastins remains unknown. Because of the predicted structure and surface localization in the intracellular stage

of *T. cruzi* and *Leishmania* spp, it has been proposed that amastins may play a role in host-parasite interactions within the mammalian cell: they could be involved in transport of ions, nutrients, across the membrane, or involved with cell signaling events that trigger parasite differentiation [9]. Its preferential expression in the intracellular stage also suggest that it may constitute a relevant antigen during parasite infection, a prediction that was confirmed by studies showing that amastins peptides elicit strong immune response during Leishmanial infection [11]. Amastin antigens are considered a relevant immune biomarker of cutaneous and visceral Leishmaniasis as well as protective antigens in mice [12].

Although complete genome sequences of two strains of *T. cruzi* (CL Brener and SylvioX-10) have been reported, their assemblies were only partially achieved because of their unusually high repeat content [13,14]. Therefore, for several multi-gene families, such as the amastin gene family, their exact number of copies is not yet known. According to the current assembly [15], only four δ -amastins and two β -amastins were identified in the CL Brener genome. Herein, we used the entire data set of sequencing reads from the CL Brener [13] and Sylvio X-10 [14] genomes, to analyzed all sequences encoding amastin orthologues present in the genomes of these two *T. cruzi* strains and determine their copy number as well as their genome organization. Expression of distinct amastin genes in fusion with the green fluorescent protein, allowed us to examine the cellular localization of different members of both amastin sub-families. By determining the levels of transcripts corresponding to each sub-family in all three parasite stages of various strains we showed that, whereas the levels of δ -amastins are up-regulated in amastigotes, β -amastin transcripts are significantly increased in the epimastigote insect stage. Most importantly, evidence indicating that amastins may constitute *T. cruzi* virulence factors was suggested by the analyses showing reduced expression of δ -amastins in amastigotes from strains known to have lower infection capacity.

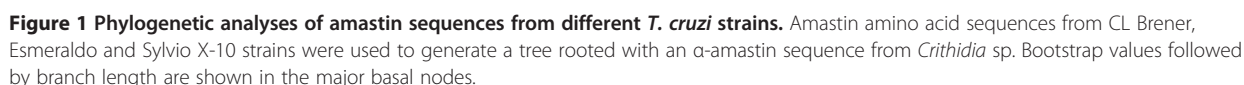
Results and discussion

The amastin gene repertoire of *Trypanosoma cruzi*

In its current assembly, the *T. cruzi* (CL Brener) genome exhibits 12 putative amastin sequences. Because of its hybrid nature and the high level of divergence between alleles, this genome was assembled as two set of contigs, each corresponding to one haplotype that were denominated Esmeraldo-like and non-Esmeraldo [13]. Therefore, the 12 amastin sequences annotated in the CL Brener genome database actually correspond to 6 pairs of alleles. Based on the analyses of amastin sequences present in the genomes of different species of *Trypanosoma* and *Leishmania*, as well as in two related insect parasites (*Leptomonas seymouri* and *Crithidia* spp.), Jackson (2010) [9] proposed a

Brener β -amastins range from 25% (between the two copies belonging to the Esmeraldo-like haplotype) and 18% (between the two non-Esmeraldo β -amastins). Analyses of additional sequences corresponding to δ -amastins, which were obtained from the individual reads generated during the CL Brener genome sequencing (see next paragraph), also show a sequence variability ranging from 85 to 100% when compared to the previously described δ -amastins. Besides the low homology found between β - and δ -amastins, low sequence identity is also found between δ -Ama40 and δ -Ama50 with the other members of the δ -amastin sub-family. On the other hand, sequence identities between members of the β -amastins or between members of the δ -amastin sub-families range from 83% up to 99% even when we compare amastins from two phylogenetically distant strains such as CL Brener and Sylvio X-10 (Additional file 1: Figure S1A).

In spite of the sequence divergence, an alignment of polypeptide sequences belonging to all amastin sub-families shows increased amino acid conservation within the putative hydrophobic transmembrane domains. Within the predicted extracellular domains, two highly conserved cysteine and one tryptophan residues, that are part of the 10 amino acid “amastin signature” [8], may be critical for amastin function (Additional file 1: Figure S1B). On the other hand, the more variable sequences present in the two predicted extracellular, hydrophilic domains suggest that this portion of the protein, which, in



amastigotes, are in contact with the host cell cytoplasm, may interact with distinct host cell proteins.

Because the assembly of CL Brener genome does not include its complete sequence, we conducted a read-based analysis to estimate the total number of amastin genes in this strain of the parasite. It is well known that the assembly of the CL Brener genome is only accurate for non-repetitive regions, and for tandemly repeated genes, misassemblies frequently occurred since most repetitive copies usually collapse into one or two copies. Therefore, we used the entire dataset of reads generated by the Tri-Tryp consortium to select reads containing sequences homologous to amastin and, based on a $13 \times$ genome coverage [13], we estimated a total number of 14 copies of amastin genes, 2 β -amastins and 12 δ -amastins in the CL Brener genome. Similar analyses performed with sequencing reads generated by Franzen *et al.* (2011) [14] from the genome of Sylvio X-10 indicated a comparable number of copies in the genome of this *T. cruzi* I strain.

In the current assembly of the CL Brener genome, amastin genes are shown to be organized in three *loci* on chromosomes 26, 32 and 34. Forty one pairs of homologous chromosomes (corresponding to the Esmeraldo-like and non-Esmeraldo haplotypes) have been assembled using the majority of the contigs and scaffolds generated by the Tri-Tryp consortium and inferences from synteny maps with the fully assembled *T. brucei* genome [15]. Based on the chromosome assemblies described by Weatherley *et al.* [15], three copies of δ -amastins are presented on chromosome 34 as a tandem array with alternating copies of tuzin genes. Interestingly, the divergent copy of δ -amastin (which has the Esmeraldo-like δ -Ama40 allele and the non-Esmeraldo allele δ -Ama50) is found as a single sequence linked to one tuzin pseudogene on chromosome 26. In a third chromosome, two copies of β -amastins are linked together without the association with tuzin genes. This gene organization is consistent with the analyses described by Jackson (2010) [9], who found tuzin genes associated only with δ -amastins. In order to confirm the proposed genomic organization in CL Brener genome and also to verify whether similar pattern of distribution of amastin genes occurs in other *T. cruzi* strains, we performed Southern blot hybridizations with chromosomal bands from CL Brener (a strain belonging to *T. cruzi* VI) as well as from G, Sylvio X-10 and Dm28c strains (all of them belonging to *T. cruzi* I) and Y strain (a *T. cruzi* II strain) separated by pulsed field gel electrophoresis. As shown in Figure 2A, the presence of two copies of β -amastins in a 900 kb chromosomal band, which is similar to the predicted size of chromosome 32 [15], has been confirmed in all *T. cruzi* strains. Using a probe specific for the δ -Ama40, we detected a chromosomal band of 800 kb, similar to the size of chromosome 26 in all strains except for the SylvioX-10, where we detected two

bands of similar sizes (Figure 2B). Since significant differences in sizes of homologous chromosomal bands in *T. cruzi* have been frequently described [16], it is possible that the two bands detected in SylvioX-10 correspond to size variation of chromosome 26 from this strain. Compared to β -amastins, the pattern of distribution of δ -amastins appears to be much more complex and variable: similar to CL Brener, in Dm28c and G strains, a probe specific for δ -amastin sub-family, which does not recognizes either β -amastins or δ -Ama40/50, hybridizes with sequences present in three chromosomal bands with approximately 1.1, 1.3 and 2.3 Mb (Figure 2C). In Sylvio X-10, Colombiana and Y strains, these sequences were found in only one or two chromosomal bands. Thus, our analyses indicates that, in addition to β -amastins, which are located in chromosome 32, members of the δ -amastin sub-family are scattered among at least 3 chromosomes in this parasite strain. Whether two of these chromosomes correspond to allelic pairs that have significant differences in size, still needs to be verified. This highly heterogeneous pattern of distribution of δ -amastin sequences is also in agreement with previous analyses described by Jackson (2010) [9], which suggest that δ -amastin sequences are apparently highly mobile. Based on analyses of genomic position as well as the phylogeny of *Leishmania* amastins, it was proposed that independent movements of δ -amastins genes occurred in the genomes of different *Leishmania* species. Also consistent with these previous analyses, when blots containing chromosomal bands were probed with a sequence encoding one of the tuzin genes, a pattern of hybridization similar to the pattern obtained with the δ -amastin probes was observed (Figure 2D). Thus, for most *T. cruzi* strains, our results are consistent with the existence of more than one cluster containing linked copies of δ -amastins and tuzin genes and an additional locus with two β -amastins linked together. However, a complete description of genomic organization of amastin genes could not be attained based solely on PFGE analyses and gene copy number estimations. Further analyses based on sequencing data generated from large inserts previously mapped on specific *T. cruzi* chromosomes are warranted to solve this question.

Distinct patterns of amastin gene expression

Because analyses of amastin gene expression have been limited to members of the δ sub-family and these studies have not been conducted with different strains of the parasite, we decided to evaluate by northern blotting the expression profiles of members of the δ - and β -amastin sub-families. We also decided to compare the expression levels of different amastin genes in parasite strains representative of *T. cruzi* I (Sylvio X-10 and G), *T. cruzi* II (Y) and in CL Brener (a *T. cruzi* VI strain). As shown in Figure 3, the levels of amastin transcripts derived from

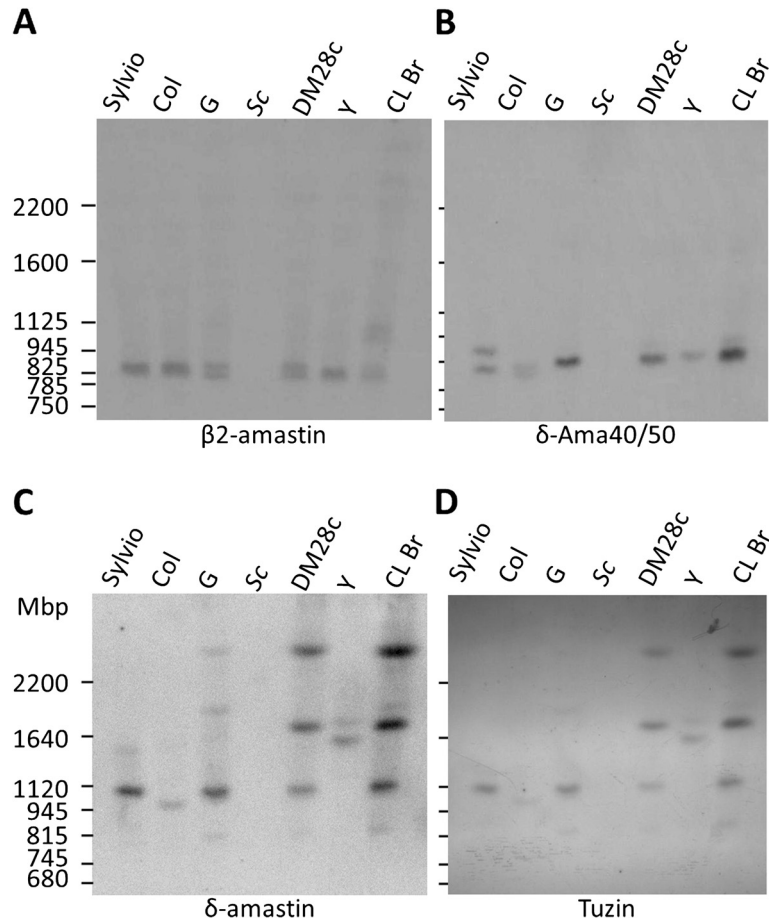


Figure 2 Genomic localization of amastin genes in different *T. cruzi* strains. Chromosomal bands from different *T. cruzi* strains, separated by Pulsed Field Gel Electrophoresis (PFGE) and transferred to membranes, were hybridized with ³²P-labelled probes corresponding to β 2-amastin (A), δ -Ama40 (B), δ -amastin (C) and tuzin genes (D). *T. cruzi* strains or clones are SylvioX-10 (Sylvio), Colombiana (Col.), G and Dm28c, Y and CL Brener (CLBr). Sizes of yeast chromosomal bands (Sc) are indicated on the left.

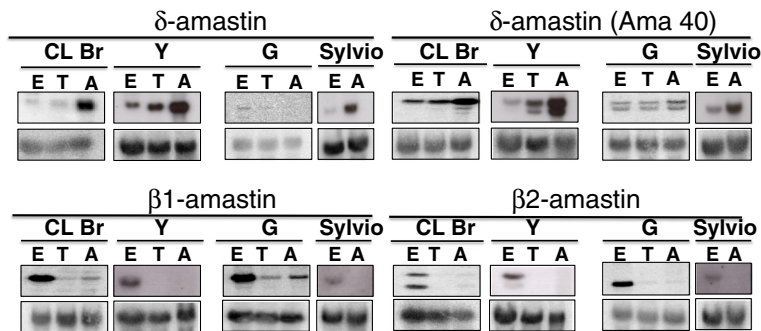


Figure 3 Amastin mRNA expression during the *T. cruzi* life cycle in different parasite strains. Total RNA was extracted from epimastigote (E), trypomastigote (T) and amastigote forms (A) from CL Brener, Y, G and Sylvio X-10. Electrophoresed RNAs (~10 μ g/lane) were transferred to nylon membranes and probed with the ³²P- labelled sequences corresponding to δ -amastin, δ -Ama40, β 1- and β 2-amastins (top panels). Bottom panels show hybridization of the same membranes with a fragment of the 24S α rRNA.

δ - and β - sub-families are differentially modulated throughout the *T. cruzi* life cycle. Most importantly, clear differences in expression levels were found when different *T. cruzi* strains are compared: whereas in CL Brener, Y and Sylvio X-10 strains, transcripts of δ -amastins are up-regulated in amastigotes, as previously described in the initial characterization of amastins performed with the Tulahuen strain (also a *T. cruzi* VI strains) [6], the same was not observed with the G strain. Even though it presents a more divergent sequence and is transcribed from a different locus in the genome, the expression of δ -Ama40, similar to other δ -amastins, is also up-regulated in amastigotes in all strains analysed except in the G strain. In contrast, in all parasite strains, the expression of β 1- and β 2-amastin transcripts is up-regulated in epimastigotes. Similar to β 2-amastin from CL Brener, two distinct δ -Ama40 transcripts with different sizes were detected in Y and G strains. It can be speculated that transcripts showing different sizes derived from δ -Ama40 and β 2-amastin genes may result from alternative mRNA processing events. Recent reports on RNA-seq analyses indicated that alternative trans-splicing and poly-adenylation as a means of regulating gene expression and creating protein diversity frequently occur in *T. brucei* [17]. Current analyses of RNA-seq data will help elucidating mechanism responsible for the size variations observed for this sub-set of β - and δ -amastins. Moreover, the striking difference in the expression of δ -amastins observed in the G strain is also currently being investigated. Because G strain has been largely characterized as a low virulence strain [18], we speculated that members of the δ -amastins sub-family may constitute virulence factors that contributed to the infection capacity and parasite survival in the mammalian host. This hypothesis has been recently verified by experiments in which we over-expressed one δ -amastin gene in the G strain and showed that the transfected parasites have accelerated amastigote differentiation into trypomastigotes in *in vitro* infections as well as parasite dissemination in tissues after infection in mice [19]. It is also noteworthy that both β -amastins exhibited increased levels in epimastigotes of all strains analysed, indicating that this amastin isoform may be involved with parasite adaptation to the insect vector. These results are consistent with previous reports describing microarray and qRT-PCR analyses of the steady-state *T. cruzi* transcriptome, in which higher levels of β -amastins were detected in epimastigotes compared to amastigotes and trypomastigote forms [20]. Similar findings were also described for one *Leishmania infantum* amastin gene (LinJ34.0730), whose transcript was detected in higher levels in promastigotes after five days in contrast to all other amastin genes that showed higher expression levels in

amastigotes [8]. The generation of knock-out parasites with the β -amastin locus deleted and pull-down assays to investigate protein interactions between the distinct *T. cruzi* amastins and host cell proteins will help elucidate the function of these proteins.

Also, to investigate the mechanisms controlling the expression of the different sub-classes of amastins, sequence alignment of the 3'UTR sequences from β - and δ -amastins were done. Previous work has identified regulatory elements in the 3' UTR of δ -amastins as well as in other *T. cruzi* genes controlling mRNA stability [4-6,21,22] and mRNA translation [23]. Since we observed that the two groups of amastin genes have highly divergent sequences in their 3'UTR (not shown), we are preparing luciferase reporter constructs to identify regulatory elements that might be present in the β -amastin transcripts as well as to identify the factors responsible for the differences observed in the amastin gene expression in distinct *T. cruzi* strains.

Amastin cellular localization

In our initial studies describing a member of the δ -amastin sub-family, we showed that this glycoprotein localizes in the plasma membrane of intracellular amastigotes [3]. Here we examine the cellular localizations of other members of the amastin family by transfecting epimastigotes of the CL Brener strain with the pTREXnGFP vector [24] containing sequences of two δ -amastins as well as β 1- and β 2-amastins in fusion with GFP. Using GFP fusion protein we were able to examine the cellular localization of each individual member of the family. Also, since several attempts of expressing the recombinant form of the full length proteins have been largely unsuccessful, it was not possible to generate specific antibodies that could be used to detect unambiguously each member of the distinct amastin sub-families. Confocal images of stably transfected epimastigotes, shown on Figure 4, demonstrated that, whereas GFP is expressed as a soluble protein present throughout the parasite cytoplasm, (Figure 4A-C) GFP fusions of β 1- and δ -amastins are clearly located at the cell surface (Figure 4D-J). Interestingly, a distinct cellular localization, with a punctuated pattern in the parasite cytoplasm of GFP fusion of δ -Ama40 as well as a more disperse distribution within the cytoplasm of the β 2- amastin GFP fusion, in addition to their surface localization was observed (Figure 4G-I and M-O). Although all amastin sequences present a N-terminal signal peptide domain, the δ -Ama40 and δ -Ama50 have a C-terminal peptide that is not present in other members of the amastin family (Additional file 2: Figure S2). In spite of these differences, all amastin sequences showed a cellular localization pattern that is consistent with the topology predicted for *Leishmania* amastins as transmembrane proteins [8], as well as with our *in silico* analyses which confirm the presence of four hydrophobic

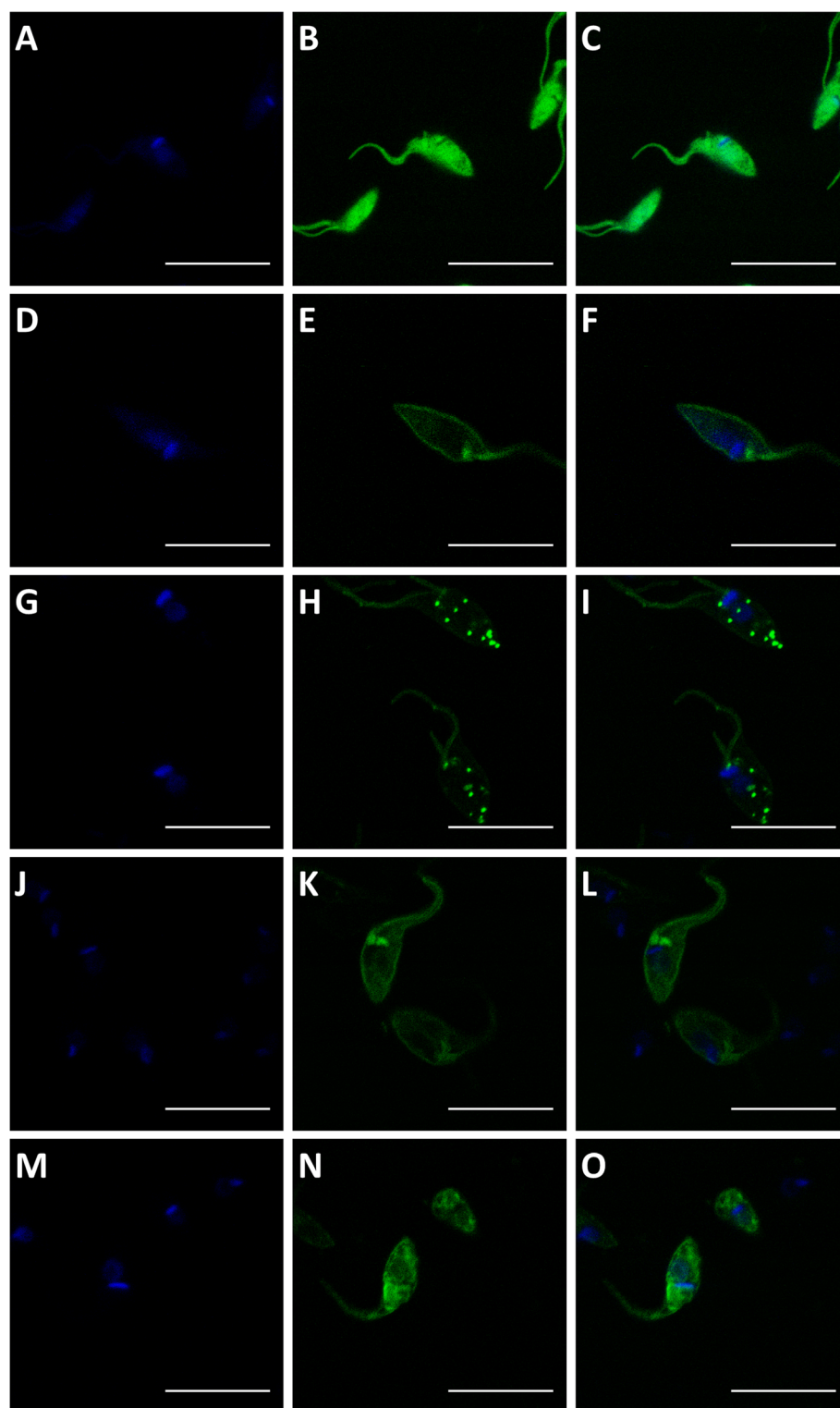


Figure 4 Subcellular localization of distinct amastins in fusion with GFP. Images from stable transfected epimastigotes of the CL Brener or G strains obtained by confocal microscopy using 1000x magnification and 2.2 digital zoom. In panels (A-C), parasites transfected with a vector containing only GFP; (D-F), parasites transfected with δ -amastinGFP; (G-I), parasites transfected with δ -Ama40GFP; (J-L), parasites transfected β 1-amastinGFP; (M-O), parasites transfected with β 2-amastinGFP. DAPI staining are shown in panels (A, D, G, J and M); GFP fluorescence in panels (B, E, H, K and N) and merged images in panels (C, F, I, L and O). (Bar = 10 μ m).

regions, a hallmark for all amastin sequences (Additional file 1: Figure S1B). To further examine their cellular localization, particularly for the δ -Ama40:GFP fusion, which may be associated with intracellular vesicles, we performed co-localization analysis with the glycosomal protein phosphoenolpyruvatecarboxykinase (PEPCK) in immunofluorescence assays. As shown by confocal images presented on Additional file 3: Figure S3, the GFP fusion protein does not co-localize with anti-PEPCK antibodies, indicating that the vesicles containing δ -Ama40 are not associated with glycosomal components. Finally, we also performed immunoblot analyses of sub-cellular fractions of the parasite and compared the presence of GFP-fusions in enriched membrane and soluble fractions of transfected epimastigotes (Figure 5). In agreement with the confocal analyses, the immunoblot results show that all four amastins that were expressed as GFP fusion proteins are presented in membrane enriched fractions.

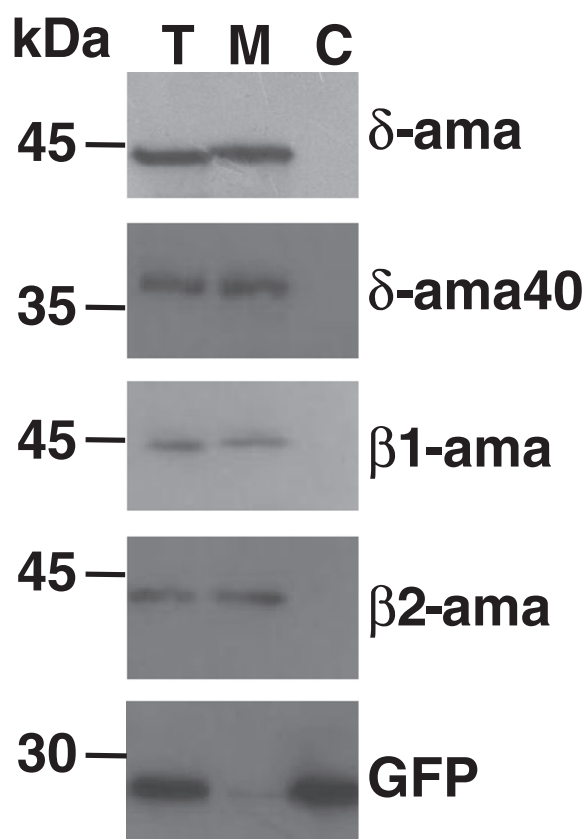


Figure 5 Distribution of amastin proteins in the parasite membrane fractions. Immunoblot of total (T), membrane (M) and cytoplasmic (C) fractions of epimastigotes expressing δ -Ama, δ -Ama40, β 1- and β 2-amastins in fusion with GFP. All membranes were incubated with α -GFP antibodies.

Conclusions

Taken together, the results present here provided further information on the amastin sequence diversity, mRNA expression and cellular localization, which may help elucidating the function of this highly regulated family of *T. cruzi* surface proteins. Our analyses showed that the number of members of this gene family is larger than what has been predicted from the analysis of the *T. cruzi* genome and actually includes members of two distinct amastin sub-families. Although most *T. cruzi* amastins have a similar surface localization, as initially described, not all amastins genes have their expression up-regulated in amastigotes: although we confirmed that transcript levels of δ -amastins are up-regulated in amastigotes from different *T. cruzi* strains, β -amastin transcripts are more abundant in epimastigotes than in amastigotes or trypomastigotes. Together with the results showing that, in the G strain, which is known to have lower infection capacity, expression of δ -amastin is down-regulated, the additional data on amastin gene expression presented here indicated that, besides a role in the intracellular, amastigote stage, *T. cruzi* amastins may also serve important functions in the insect stage of this parasite. Hence, based on this more detailed study on *T. cruzi* amastins, we should be able to test several hypotheses regarding their functions using a combination of protein interaction assays and parasite genetic manipulation.

Methods

Sequence analyses

Amastin sequences were obtained from the genome databases of *T. cruzi* CL Brener, Esmeraldo and Sylvio X-10 strains [25,26]. The sequences, listed in Additional file 4: Table S1, were named according to the genome annotation of CL Brener or the contig or scaffold ID for the Sylvio X10/1 and. All coding sequences were translated and aligned using ClustalW [27]. Amino acid sequences from CL Brener, Esmeraldo, Sylvio X-10, and *Crithidia* sp (ATCC 30255) were subjected to maximum-likelihood tree building using the SeaView version 4.4 [28] and the phylogenetic tree was built using an α -amastin from *Crithidia* sp as root. Weblogo 3.2 was used to display the levels of sequence conservation throughout the protein [29]. Amino acid sequences from one amastin from each sub-family were used to predict trans membrane domains, using SOSUI [30] as well as signal peptide, using SignalP 3.0 [31]. For copy number estimations, individual reads from the genome sequence of *T. cruzi* CL Brener [13] were aligned by reciprocal BLAST against each amastin coding sequences. Unique reads showing at least 99.7% of identity were mapped on the CDS and the coverage for each nucleotide was determined. Coverage values were normalized

through z-score and the copy numbers were determined after determining the ratios between z-score and the whole genome coverage.

Parasite culture

T. cruzi strains or clones, obtained from different sources, were classified according to the nomenclature and genotyping protocols described by [32]. Epimastigote forms of *T. cruzi* strains or clones Colombiana, G, Sylvio X-10, Dm28c, Y and CL Brener were maintained at 28°C in liver infusion tryptose (LIT) medium supplemented with 10% fetal calf serum (FCS) as previously described [3]. Tissue culture derived trypomastigotes and amastigotes were obtained after infection of LLC-MK2 or L6 cells with metacyclic trypomastigotes generated in LIT medium as previously described [3].

Pulse-field gel electrophoresis and Southern blot analyses

Genomic DNA, extracted from 10⁷ epimastigotes and included in agarose blocks were separated as chromosomal bands by pulse-field gel electrophoresis (PFGE) using the Gene Navigator System (Pharmacia) as described by Cano *et al.* (1995) [33], with the following modifications: separation was done in 0.8% agarose gels using a program with 5 phases of homogeneous pulses (north/south, east/west) with interpolation for 135 h at 83 V. Phase 1 had pulse time of 90 s (run time 30 h); phase 2 120 s (30 h); phase 3 200 s (24 h); phase 4 350 s (25 h); phase 5 800 s (26 h). Chromosomes from *Saccharomyces cerevisiae* (Bio-Rad) were used as molecular mass standards. Separated chromosomes were transferred to nylon filters and hybridized with ³²P labelled probes prepared as described in the following section.

RNA purification and Northern blot assays

Total RNA was isolated from approximately 5 × 10⁸ epimastigote, trypomastigote and amastigote forms using the RNeasy[®] kit (Qiagen) following manufacturer's

recommendations. RNA samples (15 µg/lane) were separated by denaturing agarose gel electrophoresis, transferred to Hybond-N+ membranes and hybridized with the ³²P labeled fragments corresponding to each *T. cruzi* amastin sequence as described [3]. The probes used were PCR amplified fragments from total genomic DNA extracted from the CL Brener strain using primers described in Table 1, in addition to a PCR fragment generated by amplification of the insert cloned in plasmid TcA21 (corresponding to δ-amastin) and the 24Sα ribosomal RNA [6]. DNA fragments were labeled using the Megaprime DNA-labeling kit (GE HealthCare) according to the manufacturer's protocol. All membranes were hybridized in a 50% formamide buffer for 18 h at 42°C and washed twice with 2X SSC/0.1% SDS at 42°C for 30 min each, as previously described [3]. The membranes were exposed to X-ray films (Kodak) or revealed using the STORM840 PhosphorImager (GE HealthCare).

Plasmid constructions

To express different amastin genes in fusion with GFP we initially constructed a plasmid named pTREXAmastinGFP. The coding sequence of the TcA21 cDNA clone [3] (accession number U04339) was PCR-amplified using a forward primer (5'-CATCTAGAAAGCAATGAGCAAAC-3') and a reverse primer (5'-CTGGATCCCTAGCATACGCAGAA GCAC-3') containing the *Xba*I and *Bam*HI restriction sites (underlined in the primers), respectively. After digesting the PCR product with *Xba*I and *Bam*HI, the fragment was ligated with the vector fragment of pTREX-GFP [24] that was previously cleaved with *Bam*HI and *Xho*I. To generate the GFP constructions with other amastin genes, their corresponding ORFs were PCR-amplified using the primers listed in Table 1 and total genomic DNA that was purified from epimastigote cultures of *T. cruzi* CL Brener according to previously described protocols [3]. The PCR products were cloned initially into pTZ (Qiagen) and the amastin sequences, digested with the indicated enzymes, were purified from agarose gels with Illustra GFX[™] PCR DNA and

Table 1 Sequence of primers used to amplify amastin isoforms ORFs.

Primer name / gene ID	Primer Sequence (5'-3')	Restriction enzyme
pδ1-amastin (F) Tc00.1047053511071.40	TTGTTCTAGAGTAGGAAGCAATG	<i>Xba</i> I
pδ1-amastin (R) Tc00.1047053511071.40	CGCTGGATCCGAACACGTGCA	<i>Bam</i> HI
β1-amastin (F) Tc00.1047053509965.390	CCTAGGAGGATGTCGAAGAAGAAG	<i>Avr</i> II
β1-amastin (R) Tc00.1047053509965.390	AGATCTCGAGCACAATGAGGCCAG	<i>Bgl</i> II
β2-amastin (F) Tc00.1047053509965.394	TCTAGATGGGCTTCGAAACGCTTGC	<i>Xba</i> I
β2-amastin (R) Tc00.1047053509965.394	GGATCCCCAGTGCCAGCAAGAAGACTG	<i>Bam</i> HI

The underlined sequences correspond to the restriction sites recognized by the restriction enzyme.

Gel Band Purification Kit (GE Healthcare). The fragment corresponding to the TcA21 amastin cDNA was removed from pTREXAmastinGFP after digestion with *XbaI*/*Bam*HI and the fragments corresponding to the other amastin sequences were ligated in the same vector, generating pTREXAma40GFP, pTREXAma390GFP and pTREXAma394GFP. All plasmids were purified using QIAGEN plasmid purification kits and sequenced to confirm that the amastin sequences were properly inserted, in frame with the GFP sequence.

Parasite transfections and fluorescence microscopy analyses

Epimastigotes of *T. cruzi* CL Brener, growing to a density of 1 to 2×10^7 parasites/mL, were transfected as described by DaRocha *et al.*, 2004 [24]. After electroporation, cells were recovered in 5 ml LIT plus 10% FCS 28°C for 24 h and analysed by confocal microscopy using the Confocal-Radiance2100 (BioRad) system with a 63/100x NA 1.4 oil immersion objective. To perform co-localization analyses, transfected parasites expressing amastin-GFP fusions were prepared for immunofluorescence assays by fixing the cells for 20 minutes in 4% PFA-PBS at room temperature. Parasites adhered to poly-L-lysine coverslips (Sigma) were permeabilized with 0.1% Triton X-100-PBS for 2 minutes, blocked with 4% BSA-PBS for 1 hour and incubated with primary antibodies (rabbit polyclonal antibody anti-phosphoenolpyruvate carboxykinase (anti-PEPCK, kindly provided by Stenio Fragoso, Instituto Carlos Chagas, Curitiba, Brazil) in blocking solution (5.0% non-fat dry milk) for 1 hour followed by incubation with secondary anti-rabbit IgG conjugated with Alexa546. Samples were also stained with $0.1 \mu\text{g} / \text{mL}$ 4',6-diamidino-2-phenylindole dihydrochloride (DAPI, from Sigma) at room temperature for 5 min before confocal microscopy.

Parasite membrane fractionation and western blot analyses

Aproximately 10^9 epimastigotes growing at a cell density of 2×10^7 parasites/mL were harvest, washed with saline buffer (PBS) and resuspended in lysis buffer (Hepes 20mM; KCl 10 mM; MgCl_2 1,5 mM; sacarose 250 mM; DTT 1 mM; PMSF 0,1 mM). After lysing cells with five cycles of freezing in liquid nitrogen and thawing at 37°C, an aliquot corresponding to total protein (T) extract was collected. Total cell lysate was centrifuged at a low speed ($2,000 \times g$) for 10 min and the supernatant was subjected to ultracentrifugation ($100,000 \times g$) for one hour. The resulting supernatant was collected and analysed as soluble, cytoplasmic fraction (C) whereas the pellet, corresponding to the membrane fraction (M) was resuspended in lysis buffer. Volumes corresponding to $10 \mu\text{g}$ of total parasite protein extract (T), cytoplasmic (C)

and membrane (M) fractions, mixed with Laemmli's sample buffer, were loaded onto a 12% SDS-PAGE gel, transferred to Hybond-ECLTM membranes (GE HealthCare), blocked with 5.0% non-fat dry milk and incubated with anti-GFP antibody (Santa Cruz Biotechnology) or anti-PEPCK antibody, followed by incubation with peroxidase conjugated anti-rabbit IgG and the ECL Plus reagent (GE HealthCare).

Additional files

Additional file 1: Comparative sequence analysis of *T. cruzi* amastins.

(Figure S1A) Percentages of amino acid identities among all *T. cruzi* amastin sequences present in the CL Brener and Sylvio X-10 genome databases. (Figure S1B) Conserved amino acid residues and conserved domains among sequences corresponding to all amastin genes present in the *T. cruzi* CL Brener genome are represented using the WebLogo software. The x axis depicts the amino acid position. The taller the letter the lesser the variability at the site. Predicted transmembrane domains are underlined.

Additional file 2: Amino acid sequences of delta- and beta-amastins.

(Figure S2) Predicted amino acid sequences of one representative member of δ -amastin, δ -ama40, β 1 and β 2-amastins present in the *T. cruzi* CL Brener genome.

Additional file 3: Subcellular localization of δ -Ama40 fused with GFP.

(Additional file 3: Figure S3) Permeabilized, stable transfected CL Brener epimastigotes were incubated with anti-PEPCK antibody and a secondary antibody conjugated to Alexa546. GFP (panels A and D), Alexa 546 (B and E) and merged (C and F) fluorescent images were obtained by confocal microscopy of parasites expressing δ -Ama40GFP as described in Figure 4. (Bar = $10 \mu\text{m}$).

Additional file 4: Table S1. Amastin sequences presented in Figure 1.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MMK-M, LL and WDR carried out the molecular genetic studies, microscopy analyses, sequence alignments and phylogenetic analyses. RMCP and PRA participated in molecular genetic studies. RPM-Neto and DCB participated in the sequence and phylogenetic analyses. RAM participated in the microscopy analyses. WDR and SMRT designed and coordinated the study and drafted the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

This study was supported by funds from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), Fundação de Amparo a pesquisa do Estado de Minas Gerais (FAPEMIG, Brazil) and the Instituto Nacional de Ciência e Tecnologia de Vacinas (INCTV, Brazil). DCB, RAM and SMRT are recipients of CNPq fellowships; The work of WDR, MMKM and LL is supported by Fundação Araucária, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), PPSUS/MS and CNPq.

Author details

¹Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Rua Quinze de Novembro, 1299, Centro Curitiba, PR 80060-000, Brazil.

²Departamento de Bioquímica e Imunologia, Av. Antônio Carlos, 6627, Pampulha Belo Horizonte, MG 31270-901, Brazil. ³Departamento de Parasitologia Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Pampulha Belo Horizonte, MG 31270-901, Brazil. ⁴Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, Brazil, São Paulo 04021-001, Brazil.

Received: 1 October 2012 Accepted: 14 January 2013
Published: 17 January 2013

References

- Brener Z: **Biology of *Trypanosoma cruzi*.** *Annu Rev Microbiol* 1973, **27**:347–382.
- Epting CL, Coates BM, Engman DM: **Molecular mechanisms of host cell invasion by *Trypanosoma cruzi*.** *Exp Parasitol* 2010, **126**:283–291.
- Teixeira SM, Russel DG, Kirchhoff LV, Donelson JE: **A differentially expressed gene family encoding “amastin,” a surface protein of *Trypanosoma cruzi* amastigotes.** *J Biol Chem* 1994, **269**:20509–20516.
- Coughlin BC, Teixeira SM, Kirchhoff LV, Donelson JE: **Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein.** *J Biol Chem* 2000, **275**:12051–12060.
- Araújo PR, Burle-Caldas GA, Silva-Pereira RA, Bartholomeu DC, Darocha WD, Teixeira SM: **Development of a dual reporter system to identify regulatory cis-acting elements in untranslated regions of *Trypanosoma cruzi* mRNAs.** *Parasitol Int* 2011, **60**:161–169.
- Teixeira SM, Kirchhoff LV, Donelson JE: **Post-transcriptional elements regulating expression of mRNAs from the amastin/tuzin gene cluster of *Trypanosoma cruzi*.** *J Biol Chem* 1995, **270**:22586–22594.
- Wu Y, El Fakhry Y, Sereno D, Tamar S, Papadopolou B: **A new developmentally regulated gene family in *Leishmania* amastigotes encoding a homolog of amastin surface proteins.** *Mol Biochem Parasitol* 2000, **110**:345–357.
- Rochette A, McNicoll F, Girard J, Breton M, Leblanc E, Bergeron MG, Papadopolou B: **Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp.** *Mol Biochem Parasitol* 2005, **140**:205–220.
- Jackson AP: **The evolution of amastin surface glycoproteins in Trypanosomatid parasites.** *Mol Biol Evol* 2010, **27**:33–45.
- Cerqueira GC, Bartholomeu DC, Darocha WD, Hou L, Freitas-Silva DM, Machado CR, El-Sayed NM, Teixeira SM: **Sequence diversity and evolution of multigene families in *Trypanosoma cruzi*.** *Mol Biochem Parasitol* 2008, **157**:65–72.
- Rafati S, Hassani N, Taslimi Y, Movassagh H, Rochette A, Papadopolou B: **Amastin peptide-binding antibodies as biomarkers of active human visceral *Leishmaniasis*.** *Clin Vaccine Immunol* 2006, **13**:1104–1110.
- Stober CB, Langue UG, Roberts MT, Gilmartin B, Francis R, Almeida R, Peacock CS, McCann S, Blackwell JM: **From genome to vaccines for *Leishmaniasis*: screening 100 novel vaccine candidates against murine *Leishmania* major infection.** *Vaccine* 2006, **24**:2602–2616.
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, *et al*: **The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease.** *Science* 2005, **309**:409–415.
- Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, Andersson B: **Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener.** *PLoS Negl Trop Dis* 2011, **5**:984–993.
- Weatherly DB, Boehlke C, Tarleton RL: **Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome.** *BMC Genomics* 2009, **10**:255–268.
- Souza RT, Lima FM, Barros RM, Cortez DR, Santos MF, Cordero EM, Ruiz JC, Goldenberg S, Teixeira MMG, Silveira JF: **Genome Size, Karyotype Polymorphism and Chromosomal Evolution in *Trypanosoma cruzi*.** *PLoS One* 2011, **6**:e23042.
- Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T: **Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*.** *PLoS Pathog* 2010, **6**(8):e1001037.
- Yoshida N: **Molecular basis of mammalian cell invasion by *Trypanosoma cruzi*.** *An Acad Bras Cienc* 2006, **78**:87–111.
- Cruz MC, Souza-Melo N, Vieira-da-Silva C, DaRocha WD, Bahia D, Araújo PR, Teixeira SM, Mortara RA: ***Trypanosoma cruzi*: role of delta-amastin on extracellular amastigote cell invasion and differentiation.** *PLoS One* 2012, **7**:e51804.
- Minning TA, Weatherly DB, Atwood J, Orlando R, Tarleton RL: **The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*.** *BMC Genomics* 2009, **10**:370–385.
- Araújo PR, Teixeira SM: **Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi* - A Review.** *Mem Inst Oswaldo Cruz* 2011, **106**:257–267.
- Li ZH, De Gaudenzi JG, Alvarez VE, Menciondo N, Wang H, Kissinger JC, Frasch AC, Docampo R: **A 43-nucleotide U-rich element in 3'-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes.** *J Biol Chem* 2012, **287**:19058–19069.
- McNicoll F, Müller M, Cloutier S, Boilard N, Rochette A, Dubé M, Papadopolou B: **Distinct 3'-untranslated region elements regulate stage-specific mRNA accumulation and translation in *Leishmania*.** *J Biol Chem* 2005, **280**:35238–35246.
- Darocha WD, Silva RA, Bartholomeu DC, Pires SF, Freitas JM, Macedo AM, Vazquez MP, Levin MJ, Teixeira SM: **Expression of exogenous genes in *Trypanosoma cruzi*: improving vectors and electroporation protocols.** *Parasitol Res* 2004, **92**:113–120.
- TriTryp DB: **Kinetoplastid genomic resources Database.** [http://tritypdb.org/common/downloads/release-4.1/Tcruzi/fasta/TriTrypDB].
- Aslett M, Aurecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, *et al*: **TriTrypDB: a functional genomic resource for the Trypanosomatidae.** *Nucleic Acids Res* 2010, **38**:457–462.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 20.** *Bioinformatics* 2007, **23**:2947–2948.
- Gouy M, Guindon S, Gascuel O: **SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building.** *Mol Biol Evol* 2010, **27**:221–224.
- Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**:1188–1190.
- Hirokawa T, Boon-Chieng S, Mitaku S: **SOSUI: classification and secondary structure prediction system for membrane proteins.** *Bioinformatics* 1998, **14**:378–379.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340**:783–795.
- Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, Miles MA, Romanha AJ, Sturm NR, Tibayrenc M, Schijman AG: **A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI.** *Mem Inst Oswaldo Cruz* 2009, **104**:1051–1054.
- Cano MI, Gruber A, Vazquez M, Cortés A, Levin MJ, Gonzalez A, Degraive W, Rondinelli E, Ramirez JL, Alonso C, Requena JM, Franco Da Silveira J: **Molecular karyotype of clone CL Brener chosen for the *Trypanosoma cruzi* genome project.** *Mol Biochem Parasitol* 1995, **7**:273–278.

doi:10.1186/1471-2180-13-10

Cite this article as: Kangussu-Marcolino *et al*: Distinct genomic organization, mRNA expression and cellular localization of members of two amastin sub-families present in *Trypanosoma cruzi*. *BMC Microbiology* 2013 **13**:10.